A Publication of the American Psychological Association

Journal of Abnormal and Social Psychology

Founded by Morton Prince, 1906

CONTENTS

ARTICLES

Response Styles on the MMPI: Comparison of Clinical and Normal Samples 285
Douglas N. Jackson and Samuel Messick

Anxiety, Reinforcement, and Experimental Instructions in a Free Verbalization Situation 300
Irwin G. Sarason and Victor J. Ganzer

Leader Attitudes, Group Climate, and Group Creativity 308
Fred E. Fiedler

Dissonance Reduction, Social Comparison, and Modification of Others' Opinions 319
Timothy C. Brock and James E. Blackwood

Team Competition, Success, and the Adjustment of Group Members 325
Albert Myers

Personality Style and Perceptual Reactivity to the Immediate Environment 333
Arnold Miller and Eleanor Kemp

CASE REPORT

Some Effects of Early Sensory Deprivation on Later Behavior: The Quondam Hard-of-Hearing Child 338
Nathaniel Herman Eisen

CRITIQUE AND NOTES

Identification and Empathy in Relation to Self-Dissatisfaction and Adjustment 343
Doris H. Lockwood and Bernard Guerney, Jr.

The Relationship between Response Style and Personality Variables: II. The Prediction of Independent Conformity Behavior 347
Richard K. McGee

The Instrumental Escape Conditioning of Anxious and Nonanxious Children 351
R. K. Penney and B. McCann

Volume 65 NOVEMBER 1962 Number 5

Volume 65 November 1962 Number 5

Journal of Abnormal and Social Psychology

Editor: DANIEL KATZ, *University of Michigan*

Associate Editor: ROBERT B. ZAJONC, *University of Michigan*

CONSULTING EDITORS

JOSEPH B. ADELSON
University of Michigan

GORDON W. ALLPORT
Harvard University

NORMAN H. ANDERSON
University of California, Los Angeles

LEONARD BERKOWITZ
University of Wisconsin

JACK W. BREHM
Duke University

URIE BRONFENBRENNER
Cornell University

DONALD R. BROWN
Bryn Mawr College

LAUNOR F. CARTER
System Development Corporation

DWIGHT W. CHAPMAN, JR.
Vassar College

MORTON DEUTSCH
Bell Telephone Laboratories

SEYMOUR FESHBACH
University of Pennsylvania

DONALD W. FISKE
University of Chicago

NATHANIEL L. GAGE
University of Illinois

NICHOLAS HOBBS
Peabody College for Teachers

MURRAY HORWITZ
New York University

IRVING L. JANIS
Yale University

HAROLD H. KELLEY
University of California, Los Angeles

HERBERT C. KELMAN
Harvard University

DANIEL J. LEVINSON
Harvard Medical School

DAVID C. MCCLELLAND
Harvard University

WILLIAM J. MCGUIRE
Columbia University

ROBERT B. MACLEOD
Cornell University

HAROLD L. RAUSH
National Institute of Mental Health

NEVITT SANFORD
Stanford University

IRVING SARNOFF
New York University

WILLIAM A. SCOTT
University of Colorado

JANET TAYLOR SPENCE
Veterans Administration Hospital, Iowa City

JOSEPH VEROFF
University of Michigan

JUSTIN L. WEISS
Massachusetts Mental Health Center

PUBLISHED BY

THE AMERICAN PSYCHOLOGICAL ASSOCIATION, Inc.

The JOURNAL OF ABNORMAL AND SOCIAL PSYCHOLOGY is devoted to both abnormal and social psychology. Its emphasis is on basic research and theory rather than the techniques and arts of practice. Abnormal psychology is broadly defined to include papers contributing to fundamental knowledge of the pathology, dynamics, and development of personality or individual behavior, including deterioration with age and disease. Articles concerned with psychodiagnostic techniques are evaluated with respect to their contribution to an understanding of the psychological principles of diagnosis, those concerned with psychotherapy are judged in terms of their contribution to an understanding of the therapeutic process. Case reports are considered in terms of their heuristic value. Reports which serve to pose or to clarify important theoretical problems, or which promise to be useful in teaching, are often published. From the social area this JOURNAL gives preference to papers contributing to basic knowledge of interpersonal relations, and of group influences on the pathology, dynamics, and development of individual behavior.

The JOURNAL is issued monthly, two volumes per year. The subscription price per volume is $10.00, or $20.00 annually (foreign, $20.50). Single issues are $2.00.

Original manuscripts should be sent to the Editor:

DANIEL KATZ
University of Michigan
Box 1248
Ann Arbor, Michigan

Manuscripts must be accompanied by an abstract of 100–120 words typed on a separate sheet of paper. The abstract should conform to the style of *Psychological Abstracts*. Detailed instructions for preparation of the abstracts appeared in the *American Psychologist* (1961, 16, 833), or may be obtained from the Editor or from the APA Central Office.

HELEN ORR	ELIZABETH S. REED	JUDY SMITH
Managing Editor	Advertising Manager	Editorial Assistant

Subscriptions, address changes, and business communications should be sent to:

THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.
1333 Sixteenth Street, N.W.
Washington 6, D. C.

Changes of address must reach the Subscription Office by the tenth of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee second-class forwarding postage. Other claims for undelivered copies must be made within four months of publication.

PUBLISHED BY
THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.
Prince and Lemon Sts., Lancaster, Pa.
and 1333 Sixteenth Street N.W., Washington 6, D. C.

Second-class postage paid at Lancaster, Pa., and at additional mailing offices

Copyright © 1962 by the American Psychological Association, Inc.

Journal of Abnormal and Social Psychology
1962, Vol. 65, No. 5, 285–299

RESPONSE STYLES ON THE MMPI:

COMPARISON OF CLINICAL AND NORMAL SAMPLES[1]

DOUGLAS N. JACKSON AND SAMUEL MESSICK
Pennsylvania State University *Educational Testing Service*

The investigations described in the present paper were designed: to replicate a previous study which highlighted the dominant role of the consistent response styles of acquiescence and desirability in determining response variance on the MMPI; to appraise the degree to which these stylistic consistencies are present in the responses of subjects drawn from widely differing populations; and to contrast possible differences in stylistic propensities and in consistent responses to MMPI content displayed by diverse subject groups.

In a previous study with the MMPI, based upon a prison inmate sample, Jackson and Messick (1961) utilized factor analysis to obtain a relatively precise estimation of the proportion of response variance attributable to particular response styles on the one hand and to consistent responses to item content on the other. This factor analysis of correlations among standard MMPI clinical and validity scales scored separately for true keyed and for false keyed items, along with specially constructed reference scales for acquiescence and desirability, yielded two large orthogonal factors, identified as acquiescence and the consistent tendency to respond desirably or undesirably. These two factors together accounted for over 75% of the common variance and 53% of the total variance. Of six very small factors remaining, three were interpreted as due primarily to content consistencies, and three others were considered due largely to item overlap. True and false items from the same MMPI scale, with but one exception, did not contribute to the same factors.

Because the sample of prison inmates used in the above study might be deviant and might permit only limited generalization of the results, it was considered important to ascertain the relevance of the conclusions to other populations. The present investigation therefore seeks to replicate and amplify the previous study with both a wider sampling of populations (neuropsychiatric patients and normal college students) and a broader array of MMPI scales.

Pertinent theoretical and empirical research findings have been recently reviewed by Adams (1960), Cronbach (1950), Damarin and Messick (1962), Edwards (1957, 1961), Jackson and Messick (1958), McGee (1962), Messick (1961a), Messick and Jackson (1961a), Sechrest and Jackson (1961), Stricker (1961), and Wiggins (1962).

METHOD

Subjects

Two different samples were employed, with separate analyses being performed for each. The first of these, the "hospital" sample, consisted of 194 patients—119 males and 75 females—from a state neuropsychiatric hospital. The mean age was 30.9 years, with a range from 15 to 67 and a standard deviation of 11.3.

The second group, the "college" sample, was composed of 334 undergraduates—160 males and 174 females—all paid volunteers from introductory psychology classes at a state university. The college sample was much more homogeneous in age than the hospital group, having a mean age of 19.4 years, with a range from 17 to 32 and a standard deviation of just 1.6 years.

The size of the samples would be expected to provide relatively stable correlations and factor structures and, together with results from the prison sample reported previously (Jackson & Messick, 1961), should provide a good basis for attributing generality to common findings across the three groups.

[1] This research is one of a series of studies on stylistic determinants in personality assessment supported by the National Institute of Mental Health, United States Public Health Service, under Research Grants M-2738 to Pennsylvania State University and M-2878 to Educational Testing Service.

The authors are sincerely grateful to Bernard S. Aaronson for providing the protocols for the hospital sample, to Patricia Lyons for her helpful assistance, and to Henrietta Gallagher for supervising the statistical analyses.

Procedure

Each subject in the hospital sample was administered the MMPI as part of routine admission evaluation. Patients having known intellectual deficiencies or those providing records with excessive omissions were not selected for analysis. The patients were heterogeneous with regard to psychopathological symptoms and provided good opportunities for obtaining a wide range of scores on MMPI scales.

Groups of subjects in the college sample were administered the booklet form of the MMPI under standard conditions, with assurances regarding confidential treatment of individual data.

Analysis of Desirability Scales

Separate scores for each subject were obtained for true and false keyed items of MMPI clinical and validity scales. This scoring procedure, used previously by the authors (Jackson & Messick, 1961; Messick, 1961b), permits a delineation of variance attributable to acquiescence within MMPI scales and allows for a more systematic appraisal of the effects of this variable than would be the case with the usual composite total scores obtained by adding true and false keyed items for a given scale.

Five desirability (*Dy*) scales, with all of their items keyed true, were also scored for all subjects. These five scales were developed in order to obtain scores reflecting acquiescence at systematically varying levels of item desirability (Jackson & Messick, 1958, 1961). Construction of the *Dy* scales was accomplished by dividing all MMPI items into five levels of judged desirability and using a table of random numbers to select items within each level to comprise a "scale." Item overlap between keys for *Dy* and MMPI clinical scales was systematically limited by substituting additional randomly selected items for those initial *Dy* items found to be keyed also for clinical scales. The five scales developed in this manner were labeled Dy_1 for the scale having the highest judged item desirability through Dy_5 for the scale was the lowest judged item desirability; Dy_3 was composed of items judged neutral in desirability. By obtaining *Dy* scale reliabilities, intercorrelations, and relationships with MMPI true and false keyed clinical scales, it is possible to estimate the degree of response consistency due to the generalized stylistic components of acquiescence and desirability, as contrasted with variance attributable to responses to specific item content (Jackson & Messick, 1961).

Separate "true" and "false" scores were also obtained for the following additional MMPI scales: Taylor's (1953) Manifest Anxiety (*MA*) scale; Barron's (1953) Ego Strength (Es) scale; Edwards (1957) Social Desirability (*SD*) scale; Hanley's (1957) scale of test taking (*Tt*) defensiveness; the Cofer, Chance, and Judson (1949) positive malingering (Mp) scale; and Fulkerson's (1958) acquiescence (Aq) scale. Single scores were derived for Welsh's (1956) MMPI pure factor measures, the predominantly true keyed *A* scale and the all false *R* scale.

RESULTS

Data from the hospital and college samples were analyzed separately, but since the two analyses were identical and since a comparison of the two sets of findings is of particular interest, the results from the two samples are presented in a parallel manner.

Reliabilities and Intercorrelations

Table 1 contains reliabilities (Kuder-Richardson Formula 21) and intercorrelations among the five *Dy* scales for the hospital and college samples. Reliabilities in both samples are substantial, especially for Dy_4 and Dy_5, the scales containing the least desirable items. These high reliabilities suggest that subjects responded consistently to items at a particular level of item desirability, and hence provide evidence for the presence of the stylistic tendencies elicited by the *Dy* scales. The lower reliabilities reported for Dy_3 are probably attributable both to the fact that subjects respond more consistently to item attributes other than desirability level for neutral items, and K-R Formula 21 yields a spuriously low reliability estimate for scales having a dispersion of item popularity levels, as is the case with Dy_3, for which the variance of item popularities on adult male norms (Dahlstrom & Welsh, 1960) is more than twice as great as the variances of the other four *Dy* scales.

In contrasting the reliabilities for *Dy* scales in the three samples, it is noteworthy that they are in general highest for the hospital sample, next highest for the prison sample, and lowest for the college sample. It thus appears that the group which would be expected to have the least well-organized average level of ego integration, viz., hospitalized neuropsychiatric patients, appears to show the most consistent responses to *Dy* items. There is no paradox here when it is recognized that *Dy* scales reflect response set consistencies, which might be expected to vary inversely with the degree of purposive responses to specific item content. Thus, the finding that hospital patients show the highest reliability in responding to *Dy* scales

would indicate merely that they are manifesting consistent response style effects, and are hence probably responding in a less differentiated and consistent fashion to particular item content.

Strong evidence for consistent response style variance is further indicated by an examination of the correlations among *Dy* scales in each of the two samples. Like the prison sample, both the hospital and the college samples indicate a remarkably consistent pattern of intercorrelations among *Dy* scales. In each of the samples, a particular *Dy* scale correlates most highly with other *Dy* scales similar in mean item desirability level; correlations become increasingly smaller and then negative as *Dy* scales become more dissimilar in average desirability. One way such a pattern of correlations could result would be for subjects consistently to endorse items within a certain "preferred" range of desirability and to reject items which departed in one or both directions from such a "preferred" level. Thus, a particular subject might show consistent tendencies to endorse extremely undesirable content; as content became more desirable, this tendency would become progressively less marked until the point at which item content was judged to be highly desirable, when a consistent tendency to reject such items would be in evidence. The more typical nonhospitalized subject, on the other hand, would be likely to present the reverse pattern, i.e., acquiescence to desirable and rejection of undesirable content. As a third possibility, a particular subject might show a strong tendency to acquiesce to items of neutral desirability. For such a person, items at either extreme of desirability would be weaker in acquiescence eliciting potential than items in the middle range. This type of relationship between level of item desirability and acquiescence eliciting potential is similar to Thurstone's (1959) model for attitude measurement.

TABLE 1

INTERCORRELATIONS AMONG *Dy* SCALES

Scale	Dy_1	Dy_2	Dy_3	Dy_4	Dy_5	K-R Formula 21	
						Hospital	College
Dy_1	1.00	.59	−.02	−.47	−.56	.80	.72
Dy_2	.70	1.00	.23	−.24	−.32	.71	.62
Dy_3	−.09	.23	1.00	.60	.48	.74	.58
Dy_4	−.44	−.12	.72	1.00	.81	.89	.84
Dy_5	−.65	−.34	.56	.85	1.00	.90	.79

Note.—Product-moment correlations for the hospital sample (N = 194) are below the diagonal; those for the college sample (N = 334) are above the diagonal.

TABLE 2

CORRELATIONS BETWEEN TRUE MMPI SCALES AND *Dy* SCALES: HOSPITAL AND COLLEGE SAMPLES

Desirability value		6.92 Dy_1	6.04 Dy_2	5.05 Dy_3	3.88 Dy_4	3.03 Dy_5	K-R Formula 21
SD_t	6.59	.68	.53	−.21	−.48	−.63	.47
		.59[a]	*.51*	*−.12*	*−.40*	*−.49*	*.45*
Tt_t	6.21	.32	.38	.23	.21	.02	.31
		.25	*.37*[a]	*.19*	*−.01*	*−.10*	*.18*
Mp_t	5.97	.34	.56	.40	.16	−.02	.58
		.26	*.51*[a]	*.45*	*.08*	*.03*	*.50*
Es_t	5.63	.20	.40	.52	.33	.16	.33
		.19	*.36*	*.44*[a]	*.22*	*.13*	*−.00*
Aq	4.80	−.24	.09	.78	.75	.59	.68
		−.14	*.10*	*.71*[a]	*.68*	*.49*	*.60*
Ma_t	4.20	−.32	−.02	.71	.78	.71	.72
		−.22	*.01*	*.63*	*.75*[a]	*.59*	*.60*
Si_t	3.71	−.54	−.26	.57	.82	.86	.89
		−.48	*−.39*	*.44*	*.71*[a]	*.75*	*.83*
A	3.63	−.55	−.25	.61	.83	.89	.94
		−.48	*−.34*	*.52*	*.81*	*.86*[a]	*.89*
MA_t	3.52	−.57	−.31	.54	.77	.85	.92
		−.50	*−.35*	*.47*	*.76*	*.79*[a]	*.87*
D_t	3.50	−.54	−.34	.48	.68	.80	.80
		−.45	*−.33*	*.39*	*.61*	*.69*[a]	*.56*
Pd_t	3.46	−.44	−.12	.61	.75	.75	.81
		−.44	*−.16*	*.47*	*.76*	*.74*[a]	*.69*
Pt_t	3.41	−.59	−.33	.54	.78	.89	.93
		−.40	*−.33*	*.50*	*.79*	*.81*[a]	*.85*
Hy_t	3.32	−.52	−.33	.40	.58	.72	.77
		−.44	*−.24*	*.34*	*.49*	*.55*[a]	*.39*
Sc_t	3.29	−.62	−.32	.56	.84	.92	.92
		−.53	*−.31*	*.50*	*.82*	*.82*[a]	*.83*
Hs_t	3.10	−.50	−.37	.32	.50	.62	.82
		−.27	*−.14*	*.21*	*.35*	*.45*[a]	*.57*
Pa_t	2.87	−.46	−.16	.56	.75	.81	.80
		−.44	*−.28*	*.42*	*.68*	*.67*[a]	*.61*
F_t	2.87	−.50	−.23	.51	.73	.78	.81
		−.38	*−.10*	*.45*	*.61*	*.61*[a]	*.52*

Note.—Correlations for the hospital sample (N = 194) between a given MMPI true scale and the *Dy* scales in every case appear above those for the college sample (N = 334). Correlations for the college sample are italicized.

[a] Indicates the approximate location of the highest predicted correlation between MMPI and *Dy* scales based on similar mean item desirability scale values.

Table 2 presents correlations between the *Dy* scales and true keyed items of MMPI scales for the hospital and college samples, respectively. Also given are K-R Formula 21 estimates of reliability for each measure and desirability values obtained by averaging

TABLE 3

CORRELATIONS BETWEEN FALSE MMPI SCALES AND *Dy* SCALES: HOSPITAL AND COLLEGE SAMPLES

Desirability value		6.92 Dy_1	6.04 Dy_2	5.05 Dy_3	3.88 Dy_4	3.03 Dy_5	K-R Formula 21
Ft	7.49	−.65	−.57	.02	.32	.48	.53
		[a]−.45	−.36	−.08	.22	.24	.34
Sct	6.88	−.69	−.52	.16	.44	.62	.58
		−.57[a]	−.36	.11	.36	.43	.27
Plt	6.71	−.65	−.55	.18	.44	.58	.55
		−.57[a]	−.44	.14	.43	.40	.47
Hst	6.44	−.66	−.54	.14	.37	.53	.78
		−.62	[a]−.41	.19	.43	.47	.58
MAt	6.28	−.70	−.60	.24	.45	.62	.70
		−.61	[a]−.54	.14	.43	.51	.58
Sit	5.98	−.51	−.59	−.44	−.17	.05	.73
		−.46	−.56[a]	−.42	−.11	.09	.63
Dt	5.86	−.35	−.43	−.28	−.10	.11	.52
		−.42	−.45[a]	−.33	−.10	.04	.24
Pdt	5.77	−.27	−.34	−.25	−.16	−.03	.36
		−.21	−.16	[a]−.20	−.06	.05	−.03
R	5.64	−.24	−.56	−.67	−.42	−.19	.62
		−.21	−.45	[a]−.66	−.38	−.24	.45
Hyt	5.50	−.24	−.36	−.43	−.32	−.13	.62
		−.06	−.07	[a]−.39	−.36	−.21	.33
Pat	5.04	−.03	−.18	−.51	−.55	−.36	.41
		.04	−.01	−.45[a]	−.48	−.32	.37
L	4.84	.19	.02	−.54	−.47	−.34	.60
		.12	.06	−.50[a]	−.47	−.32	.37
K	4.61	.40	.12	−.68	−.82	−.76	.76
		.35	.15	−.61	[a]−.79	−.68	.63
Mpt	4.59	.34	.08	−.65	−.71	−.60	.72
		.28	.13	−.51	[a]−.66	−.51	.58
Mat	4.53	.26	.09	−.45	−.54	−.52	.42
		.27	.29	−.14	[a]−.25	−.26	.18
Tit	4.46	.30	.06	−.60	−.71	−.58	.46
		.23	.10	−.57	[a]−.66	−.52	.34
Est	4.08	.39	.10	−.61	−.79	−.82	.80
		.34	.16	−.51	[a]−.68	−.70	.67
SDt	3.43	.63	.34	−.55	−.82	−.91	.87
		.52	.36	−.44	−.75	[a]−.79	.79

Note.—Correlations for the hospital sample ($N = 194$) between a given MMPI false scale and the *Dy* scales in every case appear above those for the college sample ($N = 334$). Correlations for the college sample are italicized.

[a] Indicates the approximate location of the highest predicted correlation between MMPI and *Dy* scales based on similar mean item desirability scale values.

desirability scale values for the items of each scale (Messick & Jackson, 1961b). The approximate location of the highest expected correlation, derived from a consideration of similar desirability values as discussed above, is indicated by an "a." Correlations for both samples are very substantial, in many cases equivalent to or higher than the attenuated K-R Formula 21 reliability estimates for these scales. The consistent patterning of these correlations indicates that each of the *Dy* scales correlates highly with MMPI scales similar to it in average level of judged desirability. Correlations for a particular MMPI scale decrease appreciably as *Dy* scales become more dissimilar in desirability, and in many cases become substantially negative. The *MA* scale, for example, is most similar to Dy_5 in desirability, and its correlation with Dy_5 is .85 in the hospital sample and .79 in the college sample. For other *Dy* scales, correlations with *MA* scale become systematically smaller as the average difference in desirability scale values increases, until with Dy_1 it becomes strongly negative— −.57 for the hospital and −.50 for the college sample. Like the previously reported data for the prison sample, it will be observed that the obtained correlations conform substantially to predictions.

Correlations between scores on false keyed MMPI scales and the five *Dy* scales (Table 3) also substantially conform to such predictions. Again, an "a" indicates the approximate location of the highest predicted correlation, based upon similar average desirability scale values. Only in this case the predicted correlation should be high and *negative*, since the false keyed MMPI scales are involved and the *Dy* scales are all keyed true. As was the case with the positive MMPI scales, the correlations are, almost without exception, highest between *Dy* and false keyed MMPI scales similar in desirability, and they systematically decrease as similarity in desirability decreases. The K-R Formula 21 estimates of reliability are lower for the negative scales, particularly in the college sample, a finding probably attributable to the shorter length of many of the false keyed scales and to the fact that in the college sample the distribution of scores was markedly skewed because of a general rejection of MMPI pathological content. Differences in the magnitudes of correlations between the three samples, however, are relatively slight when contrasted with the remarkable consistencies in the patterning of correlations within each sample.

Factor Analyses

The two matrices of intercorrelations among 40 MMPI variables (true and false parts of 15 MMPI scales, 5 *Dy* scales, and the *L*, *K*, *A*, *R*, and *Aq* scales) were factor analyzed separately for the hospital and college samples by the method of principal components. Communalities were estimated by the highest correlation in each column. An examination of the relative sizes of the 40 latent roots for each matrix led to the retention of 9 factors in the hospital sample

and 11 in the college sample, such a large number being deliberately chosen in an effort to clarify rotations (Cattell, 1958). It was also noted that there were two very large factors in each of the samples and that there was a substantial drop in the relative sizes of the latent roots from the second to the third factors. This pattern of roots was also obtained in the prison sample previously reported (Jackson & Messick, 1961), where it was found that the two dominant factors could be interpreted as acquiescence and the consistent tendency to respond desirably. The factors in each of the present groups were separately rotated analytically to an orthogonal patterned quartimax criterion (Saunders, 1960), with acquiescence and desirability as the hypothesized large factors. An additional orthogonal rotation was applied to the first two factors in each sample to balance the loadings of Dy_1 and Dy_5 and of Dy_2 and Dy_4 on the first factor, as op-

TABLE 4

FACTOR LOADINGS FOR THE HOSPITAL SAMPLE

Scale	Factor									h^2
	I	II	III	IV	V	VI	VII	VIII	IX	
1 F_t	31	71	−26	−08	06	−04	21	09	**27**	81
2 F_f	−29	64	−12	01	−16	−01	05	**29**	−20	66
3 Hs_t	07	65	−07	04	**65**	01	00	−04	02	85
4 Hs_f	−20	68	05	−06	**47**	**26**	05	10	−04	81
5 D_t	18	75	00	**47**	**25**	12	−01	−05	−02	90
6 D_f	−53	52	04	23	21	**49**	−05	04	−03	70
7 Hy_t	13	69	−01	**30**	**52**	−01	−05	03	−03	86
8 Hy_f	−61	04	04	11	**43**	**32**	−04	**28**	06	76
9 Pd_t	44	68	−01	11	−08	03	−06	19	20	77
10 Pd_f	−40	15	07	−05	−07	09	−18	**46**	06	44
11 Pa_t	39	70	−02	08	04	−06	24	07	**38**	85
12 Pa_f	−59	−23	03	17	13	18	−17	16	23	60
13 Pt_t	27	81	00	**41**	18	04	−03	−08	06	95
14 Pt_f	−16	70	−02	16	23	**35**	−02	07	−07	73
15 Sc_t	30	84	−16	17	09	−09	09	06	19	93
16 Sc_f	−16	73	−18	−09	02	07	09	**26**	02	70
17 Ma_t	60	61	−06	−07	03	−17	01	12	13	80
18 Ma_f	−34	−45	−03	−10	10	−04	15	**57**	01	68
19 Si_t	33	79	02	24	−10	02	05	**−32**	−02	93
20 Si_f	−69	29	02	08	−16	−02	−05	**−31**	00	69
21 K	−55	−68	−07	−09	06	03	03	24	13	87
22 Dy_1	30	−84	−04	03	03	−03	01	−04	−03	80
23 Dy_2	60	−61	00	05	−02	02	05	04	07	75
24 Dy_3	79	39	06	00	04	05	−10	03	03	80
25 Dy_4	58	73	−09	07	−01	00	06	−04	−02	89
26 Dy_5	30	87	−08	20	03	−01	14	−02	10	94
27 A	36	79	07	**39**	05	08	−02	−12	06	95
28 R	−81	01	−10	05	09	16	10	−21	−05	77
29 MA_t	27	79	11	**43**	16	08	−05	−11	04	94
30 MA_f	−16	76	**35**	15	**27**	07	−14	00	02	86
31 Es_t	63	−02	00	12	−19	**−39**	−20	20	−09	69
32 Es_f	−45	−64	−02	**−25**	**−28**	−07	−17	12	−17	85
33 SD_t	14	−76	**−35**	−11	−19	08	01	08	04	78
34 SD_f	−28	−85	02	**−30**	−03	−03	−06	14	−06	93
35 Tt_t	47	−18	−13	−06	00	−10	**36**	−07	−09	43
36 Tt_f	−50	−56	−05	02	04	08	07	07	**31**	69
37 Mp_t	59	−24	04	**−26**	−04	05	**34**	12	09	62
38 Mp_f	−47	−56	04	−16	−05	−01	**50**	00	−01	82
39 Ag	73	49	06	−17	−04	−05	−03	01	−06	81
40 L	−39	−39	−02	05	−00	02	**72**	03	03	83
Total Variance (in percentages)	20	39	1.2	3.7	4.2	2.2	3.4	3.3	1.6	78.6
Common Variance (in percentages)	26	49	1.6	4.7	5.3	2.8	4.4	4.2	2.0	

Note.—Loadings above .25 appear in boldface for Factors III–IX.

TABLE 5

FACTOR LOADINGS FOR THE COLLEGE SAMPLE

Scale		Factor											h^2
		I	II	III	IV	V	VI	VII	VIII	IX	X	XI	
1	F_s	35	57	−18	08	−09	−16	13	02	−17	24	00	62
2	F_t	−26	44	02	**−45**	−11	−01	00	09	−03	−01	−12	50
3	Hs_s	10	39	00	15	**27**	−06	−19	16	03	05	**−38**	47
4	Hs_t	−11	67	−08	17	−04	20	−20	**26**	**28**	−01	−03	72
5	D_s	17	69	−03	−09	**46**	18	04	−06	04	01	−13	79
6	D_t	−61	28	−01	00	08	**33**	−04	22	09	−03	06	62
7	Hy_s	14	58	−11	14	19	01	06	05	02	−01	**−45**	63
8	Hy_t	−49	−11	−17	10	09	14	−13	**55**	20	−06	−03	68
9	Pd_s	38	69	−14	−15	10	−18	06	19	−15	−05	−03	77
10	Pd_t	−26	11	05	−02	−08	00	11	**69**	−02	−05	−01	59
11	Pa_s	25	70	−12	10	22	−22	00	02	−07	−04	06	68
12	Pa_t	−50	−28	**−46**	01	11	−01	−03	12	11	−05	−09	59
13	Pt_s	31	79	−08	04	**42**	−04	−03	−08	04	−04	−03	91
14	Pt_t	−07	63	−03	01	12	**31**	−04	13	04	−11	−03	55
15	Sc_s	34	80	−15	−04	24	−13	−01	08	−11	−04	−13	88
16	Sc_t	−12	60	−09	−19	−15	03	−12	11	−11	09	**−30**	58
17	Ma_s	66	50	03	03	−09	−23	−07	04	−10	−06	−08	78
18	Ma_t	−01	−37	03	−13	−03	−07	−04	**55**	−05	17	−02	50
19	Si_s	21	76	12	**−04**	**33**	12	11	**−36**	02	09	06	91
20	Si_t	−68	32	14	−04	03	−04	13	**−27**	−02	04	04	67
21	K	−56	−65	−15	−04	−09	04	05	23	−03	04	−04	82
22	Dy_1	28	−79	07	01	11	−05	01	−04	−02	−05	06	72
23	Dy_2	50	−60	−12	−01	−02	−01	04	09	00	13	−08	65
24	Dy_3	79	31	−07	04	−01	12	09	−02	11	01	−08	77
25	Dy_4	51	76	02	−04	08	−10	−01	01	−03	−04	05	86
26	Dy_5	29	81	−05	02	22	−07	11	06	−04	13	−03	84
27	A	32	80	00	−04	**43**	04	03	−06	07	01	02	94
28	R	−80	01	−03	02	02	04	−07	−13	−04	01	19	69
29	MA_s	26	78	02	07	**43**	−03	−02	−04	17	−05	−04	91
30	MA_t	−19	71	02	11	14	02	−01	07	**52**	−08	−02	85
31	Es_s	55	−09	05	−01	−04	−15	**32**	−04	09	−11	−23	51
32	Es_t	−37	−60	10	−08	**−45**	07	10	04	−11	−06	00	75
33	SD_s	39	−68	06	09	−10	−02	−04	−02	**−55**	−04	−01	83
34	SD_t	−27	−82	−04	−12	**−32**	01	−12	12	−03	−04	08	87
35	Th_s	39	−24	−13	−02	−10	−05	**−31**	−13	−01	24	00	41
36	Th_t	−52	−53	**−33**	−01	16	02	01	12	03	00	−02	70
37	Mp_s	57	−22	−22	05	−17	07	06	01	−11	**34**	−09	59
38	Mp_t	−45	−52	08	07	−08	06	−07	−04	−09	**45**	−01	70
39	Aq	74	40	08	−02	−13	−01	−12	−01	−07	−03	05	76
40	L	−42	−32	03	−06	11	−06	−10	−01	05	**54**	01	62
Total Variance (in percentages)		18	33	1.6	1.1	4.3	1.4	1.2	4.2	2.3	2.2	1.6	70.9
Common Variance (in percentages)		26	46	2.3	1.6	6.1	2.0	1.7	6.0	3.2	3.1	2.3	

Note.—Loadings above .25 appear in boldface for Factors III–XI.

posed to the alternative position obtained by making Factor I collinear with Dy_3. The rotated factor loadings for the hospital sample are presented in Table 4 and those for the college sample in Table 5.

Plots of test vectors on the first two large dimensions for the hospital sample are presented in Figure 1. The corresponding figure for the college sample proved to be so similar that its presentation would be largely redundant. All interpretations herein based on Figure 1 are about equally appropriate when applied to the college sample.

A clear basis for the interpretation of the first factor in Figure 1 is suggested by the complete separation of true and false keyed subscales, so that in no case does a "true" scale have a negative loading or a "false" scale a positive loading. Accordingly, this factor may be interpreted relatively unequivocally as acquiescence, a consistent tendency on the part of subjects to endorse

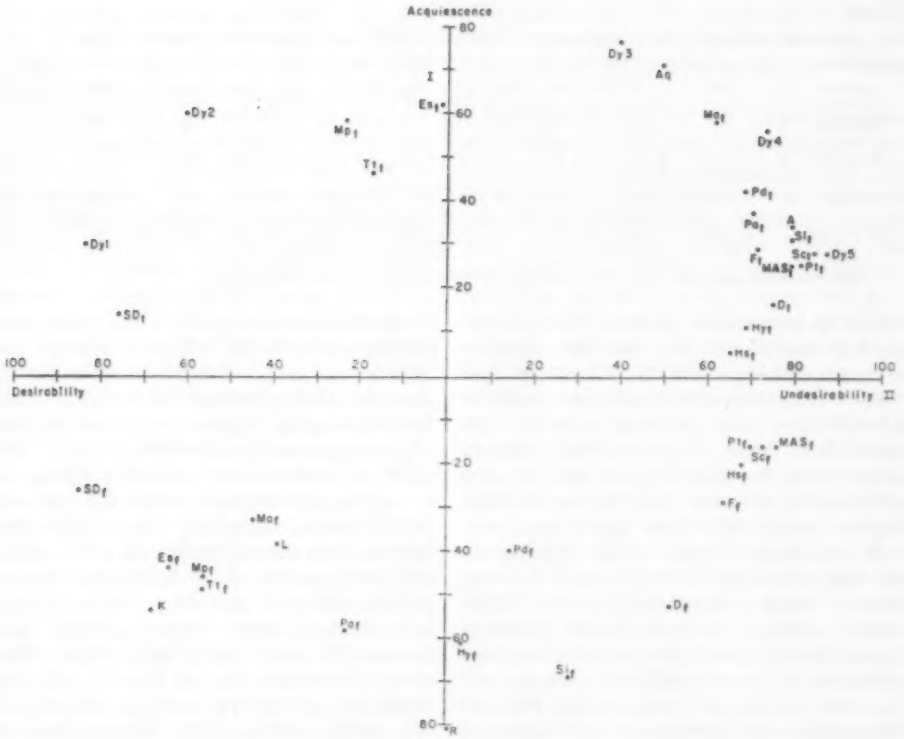

FIG. 1. Hospital sample.

or reject items. It should be noted that all of the *Dy* scales received positive loadings on this factor, a finding predictable from their all "true" keying, but that Dy_3, a scale of items neutral in desirability, received the highest loading. As scales depart from neutrality in becoming either more desirable or more undesirable, their loadings on Factor I become more moderate. Fulkerson's (1958) Acquiescence (*Aq*) scale also received a high positive loading, while Welsh's (1956) false keyed *R* scale obtained a substantial negative loading, as did the *K* scale with its predominantly "false" keying. Additional scales, such as the Taylor *MA* and Edwards *SD* scales, believed (Chapman & Campbell, 1959; and Edwards, 1961, respectively) to be relatively free from acquiescence variance in their standard form, also showed expected positive and negative loadings when true and false keyed sections were treated separately. Very few of the scales could be considered to be entirely free from acquiescence effects, a condition also illustrated by the fact that correlations between true and false parts of the same scale generally lagged behind those predicted from their respective reliabilities (Table 6).

In both the hospital and college samples, as in the prison sample previously reported, the second factor is clearly identifiable as the consistent tendency to endorse desirable or undesirable item content, as indicated by correlations above .90 in both samples between loadings on the second factor and independently obtained average desirability values for each scale (Messick & Jackson, 1961b). For both samples the scales having the largest loadings on this factor included Dy_1 in the desirable direction and Dy_5 in the undesirable, both of which had been especially designed to be extreme in desirability

TABLE 6

OBTAINED AND EXPECTED CORRELATIONS BETWEEN TRUE AND FALSE PARTS OF MMPI SCALES

Sample	Scale	*F*	*Hs*	*D*	*Hy*	*Pd*	*Pa*	*Pt*	*Sc*	*Ma*	*Si*	*MA*	*Es*	*SD*	*Tt*	*Mp*
Hospital	r_{tf}	.37	.73	.34	.18	.14	−.35	.65	.62	−.36	.16	.73	−.13	.64	−.12	.10
(N = 194)	$\sqrt{r_{tt'}r_{ff'}}$	.73	.83	.80	.81	.60	.72	.83	.81	.74	.82	.90	.77	.85	.54	.71
College	r_{tf}	.15	.35	.18	−.03	.11	−.27	.51	.48	−.16	.26	.66	−.12	.53	−.07	.05
(N = 334)	$\sqrt{r_{tt'}r_{ff'}}$	.56	.64	.70	.65	.67	.64	.71	.71	.63	.78	.87	.62	.85	.53	.64

• Reliabilities estimated here by K-R Formula 21 or the communality, whichever was higher.

but heterogeneous in content. Also obtaining high loadings in the desirable direction were the Edwards (1957) *SD* "true" and "false" subscales, which had been similarly selected for their extremeness in judged desirability. Most of the MMPI clinical scales, reflecting as they do deviant and pathological content, received quite high loadings in the undesirable direction.

A remarkable feature of these findings is the high degree of stability obtained across diverse samples. An inspection of Tables 4 and 5 reveals a strikingly similar pattern of factor loadings, even though one group was composed of neuropsychiatric patients and the other of superior young adults. The factor pattern again forms a circular array of vectors as noted previously in the prison study (Jackson & Messick, 1961), only here the circle is more complete because of the present addition of scales keyed false in the desirable direction, such as L, SD_t, Es_t, Tt_t, and Mp_t. While the means for the various clinical scales were quite different in the hospital and college groups, the underlying structure of response consistency was similar. One index of this similarity, the correlation between loadings on corresponding factors for the two samples, was in excess of .99 for the desirability dimension and was .98 for the acquiescence factor. While the average *degree* of undesirable item endorsement was higher for the hospital group, the general consistencies in responding desirably or undesirably within each group were invariant.

In both samples the first two response style factors accounted for a major proportion of the common variance. In the hospital sample, the two factors together accounted for 75% of the common variance and 59% of the total variance, and in the college sample they accounted for 72% of the common and 51% of the total variance. The communalities were also quite high, particularly in the hospital sample, frequently approaching and often surpassing the lower-bound K-R Formula 21 estimates of reliability. Thus, to a substantial degree correlations between MMPI scales, especially when subdivided into true and false subscales, could be reasonably well reproduced by knowledge of these two factors and without reference to any factors of content. These particular percentages, of course, reflect the specific method of scoring used—that of dividing standard MMPI scales into true and false keyed parts—a method which reveals the dominant role of acquiescence in influencing responses at the item level. When true and false subscales are added to yield a total score, of course, some of the acquiescence effect cancels out and the desirability factor should be even more pronounced (Edwards, 1961). However, imperfect balancing of true and false item means and variances on most MMPI scales still permits substantial generalized acquiescence effects (Messick & Jackson, 1961a).

In addition to the two large response style factors in each of the two analyses, seven factors in the hospital sample and nine in the college sample were retained for possible interpretation (Tables 4 and 5). The proportions of common and total variance accounted for by each of these factors is quite small, and an appreciable portion of this common variance is attributable to item overlap between scales, as distinguished from actual consistencies in responses to related content. The seven small factors in the hos-

pital sample together accounted for just 25% of the common variance and 19.6% of the total variance; for the nine small factors in the college sample the percentages were 28 and 19.9, respectively.

The interpretation of these factors was complicated by the problem of item overlap between scales, but response consistencies attributable to correlated content variance also appeared to underlie certain of the factors (see Jackson & Messick, 1961, for a discussion of item overlap factors and their appraisal). By eliminating from consideration factors probably attributable to item overlap, along with specific and inconsistent doublet factors, there remained four dimensions in the hospital sample (Factors IV, V, VII, VIII) and three in the college sample (Factors V, VIII, X) which seemed to warrant psychological interpretation.

Factor IV in the hospital sample and Factor V in the college sample were among the largest of the small factors. The two dimensions are quite similar—each having high loadings on true keyed neurotic and anxiety scales (D_t, MA_t, A, Pt_t, $-SD_f$, $-Es_f$, Hy_t, Hs_t, Si_t). There is considerable item overlap among these scales, but the pervasiveness of this factor and its relative size suggests that it might be interpretable in terms of content. The content of these scales reflects heterogeneous neurotic symptoms, but the overall impression suggests that "anxiety" may serve as a reasonably appropriate label. However, since all the scales load in the true direction, a more precise description would be "the tendency to endorse diverse anxiety items." A similar but broader tendency to accept a variety of psychiatric symptoms has frequently appeared as the first dimension in factor analytic studies of the MMPI (cf. Messick & Jackson, 1961a), although in the present study such a factor seems to emerge even after general acquiescence and desirability have been in effect partialed out by orthogonal axis placement. It may well be, as has been frequently suggested (Barnes, 1956a, 1956b; Jackson & Messick, 1958; Messick & Jackson, 1961a; Welsh, 1956; Wiggins, 1962), that acquiescence interacts with particular content areas in producing common patterns of responses, but such considerations should not overshadow the possible influence of item overlap in generating this particular factor.

Both Factor VII in the hospital sample and Factor X in the college sample are dominated by a high loading for the Lie scale and appear to reflect a type of naive test taking defensiveness commonly associated with elevated Lie scale scores. The interpretation of this factor is aided considerably by the inclusion of the positive malingering (Mp) scale, which was explicitly developed to assess faking. Persons scoring high on this factor would show a common tendency to respond deviantly in the desirable direction to items with a substantial disparity between their judged desirability and their modal probability of endorsement, like the item, "I read every editorial in the newspaper every day." Such items would also display large mean differences when answered under standard and faking instructions. This response tendency appears to be relatively independent of specific content areas and of the response styles of acquiescence and desirability represented in the first two factors (cf. Voas, 1958). Because responses to such scales may be interpretable in terms of both a tendency to respond consistently to certain formal properties of items and a mode of defense involving the maintenance of an excessively favorable self-concept (Marlowe & Crowne, 1961), these scales might be considered to reflect stylistic aspects of responding. It should be noted that the K scale, frequently interpreted in terms of defensiveness (Meehl & Hathaway, 1946), does not load the present factor, but is instead accounted for in terms of the first two dimensions of acquiescence and desirability.

Another dimension appearing in both the hospital and college analyses (Factor VIII in each case) was also represented in the prison sample. With the exception of Si, both parts of which load negatively, the high loadings are for false keyed scales, with MA_f and Pd_f prominently represented. Since these scales are usually associated clinically with impulsivity, alcoholism, rebelliousness, trouble with the law, employment instability,

social brashness, and the like, we previously suspected this factor to be unique to the prison inmate population, but its appearance here suggests a broader relevance.

The hospital sample yielded an additional factor (V) not obtained in the other two groups. The highest loadings on this factor were received by both true and false parts of *Hs* and *Hy*. Since this factor provides an all too infrequent instance of true and false subscales loading in the same direction, a direct content interpretation seemed warranted and the label "somatic complaints" is suggested. The failure of this factor to appear clearly in the younger, healthier, and generally more highly educated college sample is probably due to a relative lack of variance for such subjects in reporting somatic preoccupations (cf. Hollingshead & Redlich, 1958).

It is notable that none of the *Dy* scales were loaded substantially on any of the smaller factors in any of the samples. This finding suggests that the primary aim in developing these scales—that of maximizing MMPI content heterogeneity while systematically varying desirability level—was achieved.

Desirability and Communality

Wiggins (1962) has suggested that the major factor orthogonal to acquiescence on the MMPI (Factor II) be interpreted in terms of hypercommunality versus noncommunality, a dimension reflecting the tendency to endorse items in the direction of the modal response of some normative group, in contrast to responses in opposition to the modal group position. Items in between, for which approximately half the normative group responds true and half false, are said to be high in "controversiality" (Fricke, 1956; Hanley, 1957), a condition more suitable for eliciting acquiescence. This conception of communality is very similar to the interpretation of Factor II stressed in the present paper, in which the consistent endorsement of desirable versus undesirable content was emphasized. In fact, communality and desirability are so closely related that ratings of desirability have been suggested as indirect measures of item communality values and vice versa (Couch & Keniston, 1961; Wiggins, 1962; Wiggins & Rumrill, 1959)—very desirable items being expected to exhibit high proportions of endorsements and very undesirable items low proportions.

In an effort to clarify the relationship of scale communality to both the loadings on Factor II and desirability values for the scales, the average proportion of subjects answering the items in the keyed direction on adult male norms (Dahlstrom & Welsh, 1960) was computed for each scale in the present study. As expected, these communality values correlated very highly (.92) with average desirability values for the scales (Messick & Jackson, 1961b), the desirability values of the false scales being appropriately reflected to estimate the desirability of false responses (cf. Wiggins & Rumrill, 1959). The communality values were also found to be significantly correlated in the appropriate direction with loadings on the second factor —.78 in the hospital sample and .76 in the college sample. As mentioned earlier, however, the desirability values correlated .92 and .90 with these factor loadings in the hospital and college samples, respectively. Thus, over 20% more variance was associated with the factor loadings for the desirability values than for the communality values.

Since, for this particular purpose, a more optimum location of the second factor might increase its correlation with the communality values, an attempt was made to compare the maximum proportion of both communality and desirability variance predictable from dimensions optimally located in the plane of the two major factors. This was done by computing for each sample separately the two multiple correlations for predicting the communality values and the desirability values, respectively, from the loadings on the first two factors. These multiples were .88 and .86 for predicting the communality values in the hospital and college samples, respectively; whereas the corresponding multiples for predicting the desirability values were .96 and .95.

Thus, communality and desirability are closely related to each other and both are

substantially associated with the second factor, which in turn should probably be interpreted as a function of both. However, since slightly more variance was involved for the desirability values than for the communality values, moderate support is given to the present preference for weighting desirability somewhat more heavily than communality in interpreting the second factor.

DISCUSSION

The results from the present study based upon hospital and college samples, when considered together with those from the prison sample previously reported, reveal a strikingly consistent pattern. In all three analyses, approximately three-fourths of the common variance and about one half of the total variance was interpretable in terms of the response styles of acquiescence and desirability. Of the remaining very small factors, roughly half their variance was attributable to item overlap, and hence is of little or no psychological significance. While consistencies interpretable in terms of reliable responses it item content did indeed appear in all three samples, with moderate replication in the diverse groups, their aggregate contribution to the total response variance was not great in comparison with other sources of variance. To interpret scale scores presumably reflecting such content factors without taking into account the rather massive response style effects would be quite hazardous.

While our procedure of dividing standard scales into true and false subscales admittedly tended to highlight acquiescence in the response variance of the MMPI, it also revealed another important property of MMPI scales. Even after the general acquiescence factor was removed in the three samples, true and false keys for a given MMPI scale showed a marked tendency *not* to load on the same factor. In the prison sample only the *D* scale received consistent loadings for true and false parts on the same content factor. Just two of the scales, *Si* and *Mf*, yielded consistent loadings for true and false items in both the college and hospital samples; whereas *Hs* and *Hy* received consistent loadings on a factor found in the hospital sample only. In addition to the questions raised by this unfortunate property of most MMPI scales with respect to the logic of adding subscales which fail to share common variance, this finding does not add to the confidence with which characteristics can be unequivocally attributed to individuals on the basis of total scores obtained by adding true and false items for particular MMPI scales.

Although it was assumed that placing the first two response style dimensions orthogonal to the rest would remove from the remaining factors the massive general effects of acquiescence and desirability (or at least their linear components), it still seems that consistent tendencies to respond true, in interaction with particular content, continue to operate in determining response variance. This is particularly notable in Factor IV of the hospital sample and Factor V of the college sample, for which a true response to items representing a variety of neurotic contents seems to be the common denominator. There is no correspondingly large factor involving the false keyed items of these scales. It may well be that true items are more "valid" than false items for assessing certain content areas (cf. Gage & Chatterjee, 1960), and the present results suggest that this might be true even after the major effect of generalized acquiescence is removed. A similar tendency may be noted in Factor VIII of both the hospital and college samples, which is characterized by a differential tendency to endorse false keyed items generally considered to reflect impulsivity, social brashness, and lack of control and socialization, among other things. The fact that both true and false parts of *Si* received appropriate loadings, however, suggests that it would be possible to write other items to define both extremes of this dimension.

Another noteworthy finding is the fact that relationships among the *Dy* scales were not entirely symmetrical, even though their average desirability values were fairly evenly spaced. In Table 1 the correlations of Dy_3 with the remaining *Dy* scales indicate a tendency for acquiescence to items moderate or neutral in desirability to be more

highly associated with endorsement of undesirable rather than desirable content. In both samples the correlation of Dy_3 with Dy_1 was negative, whereas with Dy_5 it was substantially positive. This finding cannot be accounted for by differences in the average desirability values of the scales, since if anything, Dy_3 is closer to Dy_1 than Dy_5 in average desirability. Consistent with this result is the fact that both measures of acquiescence used in the present study—Dy_3 and Aq—received moderate loadings in the undesirable direction of Factor II (Tables 4 and 5). Another way to indicate this relationship, then, would have been to rotate the factors obliquely, placing Factor I between Dy_3 and Aq, to obtain a moderate correlation between the two factors of acquiescence and undesirability. These findings, which are consistent with Jackson's (1960) results with the California Psychological Inventory, suggest that if the tendency to endorse heterogeneous items moderate in desirability can be taken as a criterion measure of at least one type of acquiescence, then persons displaying such acquiescent tendencies may be somewhat more likely to endorse undesirable rather than desirable content. This suggestion is in conformity with accumulating evidence linking acquiescence to authoritarianism (Gage & Chatterjee, 1960) low intelligence (Messick & Frederiksen, 1958), rigidity (Jackson, Messick, & Solley, 1957), anxiety (Cattell & Scheier, 1961), and other characteristics usually considered unfavorable (cf. Couch & Keniston, 1960).

The possibility of a substantive relationship between acquiescence and the tendency to respond undesirably also offers an alternate basis for resolving the recent controversy over the confounding of desirability variance in a measure of acquiescence (Edwards & Walker, 1961a, 1961b) and the countercharge of having acquiescence involved in a presumed measure of desirability (Couch & Keniston, 1961). The present results are particularly relevant to this issue, since the pattern of relationships among several scales was used as the criterion of rotation and not the location of any particular acquiescence or desirability scale. From the present view, neither the finding of a negative loading for an acquiescence scale on a desirability factor (Edwards & Walker, 1961b) nor the finding of a negative loading for *SD* on an acquiescence factor (Couch & Keniston, 1960) should *necessarily* be interpreted as a derogation of the scale properties of either measure—the loadings may reflect a genuine psychological relationship. Both of the scales in question could probably benefit from purification, but item selection to eliminate the correlation between them should not serve as a criterion of goodness.

Turning now to the more general implications of our results, we should like to make a few observations about the present status of the assessment of psychopathology with structured inventories. Recent research on the internal structure of inventories like the MMPI has consistently revealed systematic, cumulative, and very substantial response style effects which not only cannot be ignored, but which raise important questions as to the conditions, if any, under which responses to structured items can be interpreted in terms of some particular content or criterion relevant to psychopathology. The response styles themselves, of course, may be related to psychopathology, and the search for identifiable personality correlates of these styles (Jackson & Messick, 1958) should certainly continue, for there are already some promising initial results (Bass, 1956; Cattell, 1957; Clayton & Jackson, 1961; Couch & Keniston, 1960; Damarin & Messick, 1962; Frederiksen & Messick, 1959; Jackson, 1959; Jackson et al., 1957; Jackson & Pacine, 1961; McGee, 1962; Marlowe & Crowne, 1961; Messick & Frederiksen, 1958). However, with a true-false item format as in the MMPI, one rapidly approaches an upper limit on the amount of information elicited by items which permit massive response style effects like generalized acquiescence and desirability bias. With dichotomous items yielding only "deviant-true" and "deviant-false" responses, one can identify respondents who vary predominantly on only two general stylistic dimensions (Jackson & Messick, 1958; Sechrest & Jackson, 1961; Wiggins, 1962). If these two dimensions account for a very

substantial proportion of the response variance, then the usefulness of the available information for differentiating among respondents with respect to particular psychopathological states or conditions must necessarily be limited. It is true that the presence or absence of various psychopathological conditions might well be reflected in differing weights for the two major response styles, a possibility consistent with several attempts to differentiate among clinical groups psychometrically which have indeed uncovered only two canonical variates (Beech & Maxwell, 1958; Eysenck & Claridge, 1962; McCarter, 1961; Rao & Slater, 1949). However, the empirical fact that various psychopathological states yield stylistic response patterns on the MMPI which are highly intercorrelated provides a basis for understanding the general lack of discriminant validity associated with MMPI scale scores. To be sure, there are a few other potentially general response styles which might be elicited from MMPI items, such as the differential tendency to endorse deviantly items having high desirability but low communality values, like the Lie scale (Wiggins, 1962), or to prefer items with absolutely or probabilistically phrased qualifiers (Clayton & Jackson, 1961). However, it is doubtful that the number of such potential response style variables is high on the MMPI, nor is their relevance to psychopathology known with certainty.

There is a question then as to whether or not the item format of the MMPI, providing such fertile ground as it does for eliciting powerful acquiescence and desirability effects, is the method of choice in the structured assessment of psychopathology. This question naturally devolves around issues of validity per unit cost (Cronbach & Gleser, 1957). Berg (1955, 1959) has already suggested that it might be more economical to use a small number of items dealing with preferences for abstract drawings—provided one's purpose is to differentiate schizophrenics from normals (Adams & Berg, 1961a, 1961b; Barnes, 1955). However, in addition to convergent validity, there is a need for discriminant validity (Campbell & Fiske, 1959). We submit that typically the clinician is at least as interested in differentiating a schizophrenic from a depressive as he is in distinguishing hospitalized patients from normals. While massive response sets apparently contribute to the differentiation of normals from mental patients and thus to convergent validity, in most cases such sets will prove to be a definite hinderance to unequivocal and refined differential diagnosis, and therefore militate against discriminant validity. Therefore, where interest is centered upon the differential assignment of characteristics relevant to personality or to psychopathology on the basis of test scores, experimental controls for response set biases are imperative. The particular form that such controls should take in the assessment of psychopathology is a question for further research (Adams, 1960; Messick, 1960, 1961b; Messick & Jackson, 1958, 1961a).

SUMMARY

In an effort to appraise the degree to which response consistencies on the MMPI are attributable in widely differing populations to the general sytlistic dimensions of acquiescence and desirability on the one hand, and to particular content consistencies on the other, two studies on two diverse samples were performed. By applying factor analysis to correlations among MMPI clinical and validity scales scored separately for true and false items, along with five specially constructed reference scales to measure tendencies to endorse items at varying levels of desirability, relatively precise estimates of the proportion of variance attributable to response style and to content were obtained. The two samples employed—hospitalized neuropsychiatric patients and college students—yielded results consistent with those obtained from a previous study of prison inmates, in that the dominant role of acquiescence and desirability response styles was emphasized. In all three samples two very large factors appeared, identifiable as acquiescence and desirability. On the average, approximately three-quarters of the common variance and over half of the total variance was attributable to the two stylistic dimensions, involving in the hospital sample a

somewhat larger and in the college sample a somewhat smaller proportion of the variance as compared with the previous prison sample. All MMPI subscales employed in these three analyses elicited substantial response style effects. Furthermore, in each of the three analyses the factor loadings of the scales on the acquiescence and desirability dimensions arrayed the scale vectors in a circular pattern, suggesting that acquiescence is differentially elicited by items at varying levels of desirability.

In addition, several quite small factors were obtained in each of the analyses, some attributed to item overlap and some to consistent content responses. Four factors in the hospital sample and three in the college sample were identified tentatively in terms of item content. Three factors more or less common to the two groups were interpreted, respectively, as: endorsement of items reflecting anxiety symptoms, naive test taking defensiveness, and social brashness and impulsivity. One factor, appearing only in the hospital sample, was identified as reflecting somatic complaints.

The massive response style effects on the MMPI, while contributing to convergent validity in grossly separating normals from mental patients, are interpreted as reducing the possibilities for discriminant validity, as in the differential diagnosis of psychopathological syndromes. Where the latter is important, experimental controls aimed at sharply reducing response biases are recommended.

REFERENCES

ADAMS, GEORGIA S. Techniques of minimizing or capitalizing upon response tendencies in structured self-report inventories. Paper read at International Congress of Psychology, Bonn, Germany, August 1960.

ADAMS, H. E., & BERG, I. A. Affective tone of test option choice as a deviant response. *Psychol. Rep.*, 1961, **8**, 79–85. (a)

ADAMS, H. E., & BERG, I. A. Schizophrenia and deviant response sets produced by auditory and visual test content. *J. Psychol.*, 1961, **51**, 393–398. (b)

BARNES, E. H. The relationship of biased test responses to psychopathology. *J. abnorm. soc. Psychol.*, 1955, **51**, 286–290.

BARNES, E. H. Factors, response bias, and the MMPI. *J. consult. Psychol.*, 1956, **20**, 419–421. (a)

BARNES, E. H. Response bias and the MMPI. *J. consult. Psychol.*, 1956, **20**, 371–374. (b)

BARRON, F. An ego strength scale which predicts response to psychotherapy. *J. consult. Psychol.*, 1953, **17**, 327–333.

BASS, B. M. Development and evaluation of a scale for measuring social acquiescence. *J. abnorm. soc. Psychol.*, 1956, **53**, 296–299.

BEECH, H. R., & MAXWELL, A. E. Differentiation of clinical groups using canonical variates. *J. consult. Psychol.*, 1958, **22**, 113–121.

BERG, I. A. Response bias and personality: The deviation hypothesis. *J. Psychol.*, 1955, **40**, 60–71.

BERG, I. A. The unimportance of test item content. In B. M. Bass & I. A. Berg (Eds.), *Objective approaches to personality assessment*. Princeton, N. J.: Van Nostrand, 1959. Pp. 83–99.

CAMPBELL, D. T., & FISKE, D. W. Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychol. Bull.*, 1959, **56**, 81–105.

CATTELL, R. B. *Personality and motivation structure and measurement*. Yonkers-on-Hudson, N. Y.: World Book, 1957.

CATTELL, R. B. Extracting the correct number of factors in factor analysis. *Educ. psychol. Measmt.*, 1958, **18**, 791–838.

CATTELL, R. B., & SCHEIER, I. H. *The meaning and measurement of neuroticism and anxiety*. New York: Ronald, 1961.

CHAPMAN, L. J., & CAMPBELL, D. T. Absence of acquiescence response set in the Taylor manifest anxiety scale. *J. consult. Psychol.*, 1959, **23**, 465–466.

CLAYTON, MARTHA B., & JACKSON, D. N. Equivalence range, acquiescence, and overgeneralization. *Educ. psychol. Measmt.*, 1961, **21**, 371–381.

COFER, C. N., CHANCE, J. E., & JUDSON, A. J. A study of malingering on the Minnesota Multiphasic Personality Inventory. *J. Psychol.*, 1949, **27**, 491–499.

COUCH, A., & KENISTON, K. Yeasayers and naysayers: Agreeing response set as a personality variable. *J. abnorm. soc. Psychol.*, 1960, **60**, 151–174.

COUCH, A., & KENISTON, K. Agreeing response set and social desirability. *J. abnorm. soc. Psychol.*, 1961, **62**, 175–179.

CRONBACH, L. J. Further evidence on response sets and test design. *Educ. psychol. Measmt.*, 1950, **10**, 3–31.

CRONBACH, L. J., & GLESER, GOLDINE C. *Psychological tests and personnel decisions*. Urbana: Univer. Illinois Press, 1957.

DAHLSTROM, W. G., & WELSH, G. S. *An MMPI handbook: A guide to use in clinical practice and research*. Minneapolis: Univer. Minnesota Press, 1960.

DAMARIN, F. L., & MESSICK, S. Response styles as personality variables: Evidence from published multivariate research. Research Bulletin, Educational Testing Service, 1962, in press.

EDWARDS, A. L. *The social desirability variable in personality assessment and research.* New York: Dryden, 1957.

EDWARDS, A. L. Social desirability or acquiescence in the MMPI? A case study with the *SD* scale. *J. abnorm. soc. Psychol.*, 1961, **63**, 351–359.

EDWARDS, A. L., & WALKER, J. N. A note on the Couch and Keniston measure of agreeing response set. *J. abnorm. soc. Psychol.*, 1961, **62**, 173–174. (a)

EDWARDS, A. L., & WALKER, J. N. Social desirability and agreement response set. *J. abnorm. soc. Psychol.*, 1961, **62**, 180–183. (b)

EYSENCK, H. J., & CLARIDGE, G. The position of hysterics and dysthymics in a two-dimensional framework of personality description. *J. abnorm. soc. Psychol.*, 1962, **64**, 46–55.

FREDERIKSEN, N., & MESSICK, S. Response set as a measure of personality. *Educ. psychol. Measmt.*, 1959, **19**, 137–159.

FRICKE, B. G. Response set as a suppressor variable in the OAIS and MMPI. *J. consult. Psychol.*, 1956, **20**, 161–169.

FULKERSON, S. C. An acquiescence key for the MMPI. *USAF Sch. Aviat. Med. Rep.*, July 1958, No. 58-71.

GAGE, N. L., & CHATTERJEE, B. B. The psychological meaning of acquiescence set: Further evidence. *J. abnorm. soc. Psychol.*, 1960, **60**, 280–283.

HANLEY, C. Deriving a measure of test-taking defensiveness. *J. consult. Psychol.*, 1957, **21**, 391–397.

HOLLINGSHEAD, A. B., & REDLICH, F. C. *Social class and mental illness: A community study.* New York: Wiley, 1958.

JACKSON, D. N. Cognitive energy level, acquiescence, and authoritarianism. *J. soc. Psychol.*, 1959, **49**, 65–69.

JACKSON, D. N. Stylistic response determinants in the California Psychological Inventory. *Educ. psychol. Measmt.*, 1960, **20**, 339–346.

JACKSON, D. N., & MESSICK, S. Content and style in personality assessment. *Psychol. Bull.*, 1958, **55**, 243–252.

JACKSON, D. N., & MESSICK, S. Acquiescence and desirability as response determinants on the MMPI. *Educ. psychol. Measmt.*, 1961, **21**, 771–792.

JACKSON, D. N., MESSICK, S., & SOLLEY, C. M. How rigid is the authoritarian? *J. abnorm. soc. Psychol.*, 1957, **54**, 137–140.

JACKSON, D. N., & PACINE, L. Response styles and academic achievement. *Educ. psychol. Measmt.*, 1961, **21**, 1015–1028.

MCCARTER, R. E. Emotional components of early recollections. Technical Report, 1961, Educational Testing Service, New Jersey.

MCGEE, R. K. Response style as a personality variable: By what criterion? *Psychol. Bull.*, 1962, **59**, 284–295.

MARLOWE, D., & CROWNE, D. P. Social desirability and response to perceived situational demands. *J. consult. Psychol.*, 1961, **25**, 109–115.

MEEHL, P. E., & HATHAWAY, S. R. The *K* factor as a suppressor variable in the MMPI. *J. appl. Psychol.*, 1946, **30**, 525–564.

MESSICK, S. Dimensions of social desirability. *J. consult. Psychol.*, 1960, **24**, 279–287.

MESSICK, S. Personality structure. *Annu. Rev. Psychol.*, 1961, **12**, 93–128. (a)

MESSICK, S. Response style and content measures from personality inventories. *Educ. psychol. Measmt.*, 1961, **21**, 915–924. (b)

MESSICK, S., & FREDERIKSEN, N. Ability, acquiescence, and "authoritarianism." *Psychol. Rep.*, 1958, **4**, 687–697.

MESSICK, S., & JACKSON, D. N. The measurement of authoritarian attitudes. *Educ. psychol. Measmt.*, 1958, **18**, 241–253.

MESSICK, S., & JACKSON, D. N. Acquiescence and the factorial interpretation of the MMPI. *Psychol. Bull.*, 1961, **58**, 299–304. (a)

MESSICK, S., & JACKSON, D. N. Desirability scale values and dispersions for MMPI items. *Psychol. Rep.*, 1961, **8**, 409–414. (b)

RAO, C. R., & SLATER, P. Multivariate analysis applied to differences between neurotic groups. *Brit. J. Psychol., Statist. Sect.*, 1949, **2**, 17–29.

SAUNDERS, D. R. A computer program to find the best-fitting orthogonal factors for a given hypothesis. *Psychometrika*, 1960, **25**, 199–205.

SECHREST, L. B., & JACKSON, D. N. Deviant response tendencies: Their measurement and interpretation. Research Bulletin No. 19, 1961, Pennsylvania State University.

STRICKER, L. Some item characteristics that evoke acquiescent and social desirability response sets on psychological scales. Research Bulletin, 1961, Educational Testing Service, New Jersey.

TAYLOR, JANET A. A personality scale of manifest anxiety. *J. abnorm. soc. Psychol.*, 1953, **48**, 285–290.

THURSTONE, L. L. *The measurement of values.* Chicago: Univer. Chicago Press, 1959.

VOAS, R. B. Relationships among three types of response sets. *USN Sch. Aviat. Med. proj. Rep.*, No. 15, 1958, Proj. No. NM 16 01 11, Sub. 1, No. 15.

WELSH, G. S. Factor dimensions *A* and *R*. In G. S. Welsh & W. G. Dahlstrom (Eds.), *Basic readings on the MMPI in psychology and medicine.* Minneapolis: Univer. Minnesota Press, 1956. Pp. 264–281.

WIGGINS, J. S. Strategic, method, and stylistic variance in the MMPI. *Psychol. Bull.*, 1962, **59**, 224–242.

WIGGINS, J. S., & RUMRILL, C. Social desirability in the MMPI and Welsh's factor scales *A* and *R*. *J. consult. Psychol.*, 1959, **23**, 100–106.

(Received October 5, 1961)

Journal of Abnormal and Social Psychology
1962, Vol. 65, No. 5, 300–307

ANXIETY, REINFORCEMENT, AND EXPERIMENTAL INSTRUCTIONS IN A FREE VERBALIZATION SITUATION [1]

IRWIN G. SARASON AND VICTOR J. GANZER
University of Washington

The studies of verbal conditioning carried out in recent years have provided good evidence that the subject's usage of many different kinds of response classes can be increased as a function of reinforcement. For example, it has been shown that the experimenter's reinforcement of plural nouns, in an experimental situation, leads to an increase in the subject's usage of plural nouns (Krasner, 1958). There are many interesting possibilities suggested by this type of finding. One of these is the extension of verbal conditioning procedures to situations which more resemble free verbalization and interview situations than do those heretofore studied in most verbal conditioning experiments. The present report describes one experimental attempt at such an extension.

Each subject in this experiment was asked to present a verbal description of himself during a 40-minute period. Following a 10-minute operant period, each subject was treated in one of three ways. One-third of the subjects were reinforced for making positive self-references, one-third were reinforced for making negative self-references, and one-third were given no reinforcements. Rogers (1960), using a situation similar to the one used in the present study, has reported a conditioning effect for subjects reinforced for negative self-references. The effect of reinforcement of positive self-references seemed to be that of arresting extinction.

The aim of the present research was to relate one personality variable, test anxiety, and one experimental variable, preliminary instructions, to the subject's self-descriptions during the operant period, and to the changes in the subject's self-descriptions under reinforcement conditions. Each subject received one of two sets of preliminary instructions. Labeled as the Threat condition were instructions which informed the subject that the experimental situation would provide clinical information about personality deviations. Labeled as the Nonthreat condition were instructions which informed the subject that the purpose of the experiment was simply to increase knowledge of college students' attitudes and feelings.

On the basis of theoretical considerations and available empirical evidence, certain hypotheses concerning the effects of these manipulations were made. On the assumption that score on the test anxiety scale reflects a tendency to see oneself or, at least, to describe oneself as an anxious individual, low in feelings of self-adequacy, it was hypothesized that during the operant period high anxious (HA) subjects would emit a higher percentage of negative self-references than would low anxious (LA) subjects. It was also assumed that HA individuals differ from LA individuals in their greater tendency to respond to threatening situations with feelings of self-derogation. On this basis, as well as on the basis of evidence which indicates that HA subjects may be more responsive to the experimenter's reinforcements in verbal conditioning experiments (Sarason, 1960), it was hypothesized that HA subjects would show a greater increase than would LA subjects in response classes and, particularly, in negative self-references following their reinforcement under the Threat than under the Nonthreat condition. Because the threat situation involved an evaluational emphasis, the measure of anxiety selected was the Test Anxiety Scale, a scale dealing with the subject's reactions to test situations.

The results of this experiment, it was felt, would be of value with respect to two related general problems: one, the analysis of vari-

[1] This research was supported by State of Washington Initiative 171 funds for Research in Biology and Medicine.

ables and procedures involved in changes in verbal behavior; the other, the isolation of behavioral correlates of anxiety.

METHOD

Subjects. The subjects were 48 male and 48 female freshman and sophomore students enrolled in introductory psychology classes at the University of Washington. Males and females were subdivided with respect to test anxiety scores, in terms of the upper and lower 20% of the score distribution. High test anxiety scale scores were 9 and above; low scores were 2 and below. Table 1 lists the 16 items in the Test Anxiety Scale (TAS).

Procedure. A $7\frac{1}{2}' \times 12\frac{1}{2}'$ experimental room was equipped with a Wollensack Model T-1500 tape recorder with a built-in signaling device, a $2\frac{1}{2}' \times 4'$ table, and two chairs. The recorder was on a low box beside the experimenter's chair and not in view from the opposite side of the table.

The signaling device consisted of an insulated wire soldered inside the microphone jack on the same contact used by the microphone cord itself. The free end of the wire was connected to a doorbell button attached to the underside of the table. When the experimenter wished to superimpose a tone over the tape recording, he depressed the button with his knee, thereby introducing a buzz into the recording which was inaudible to both the subject and the experimenter. This signal tone was used to mark the end of each 10-minute interview period, and as a signal given simultaneously with the experimenter's verbal reinforcements.

Directly behind the experimenter's chair was a portable $5' \times 7'$ beaverboard partition extending the width of the room. This panel was utilized as an acoustical control measure to dampen vocal vibrations and facilitate clearer tape recordings.

In the experimental situation, the subject and the experimenter were seated facing each other across the table. Half of the high and half of the low TAS subjects received the following instructions, labeled as the Threat condition:

> The interview is an excellent method for assessing a person's thinking and behavior patterns. This interview is a technique for discovering conscious and unconscious personality maladjustments and neurotic tendencies. From the statements students make about themselves, it is possible to discover certain important characteristics related to potential mental health problems. I want to find out how students think and feel about *themselves*. I'd like to know what you think you are like as a person. I believe the best way to accomplish this is simply to ask you to talk about and describe yourself.
>
> In this experiment, your instructions are to describe yourself, and what you think about yourself and how you feel about yourself. Naturally, I'll say nothing to interrupt you since I'm interested in listening to *you* talk. We'll begin now with your description. You'll have a little over half an hour for this. I'll tell you when to stop. Do you have any questions?
>
> Please hold this microphone. Please try not to tap it on the table, as it is very sensitive and will pick up unwanted noise. OK? Let's begin.

TABLE 1

TEST ANXIETY SCALE ITEMS

Item
1. While taking an important examination, I perspire a great deal. (T)
2. I get to feel very panicky when I have to take a surprise exam. (T)
3. During tests, I find myself thinking of the consequences of failing. (T)
4. After important tests I am frequently so tense that my stomach gets upset. (T)
5. While taking an important exam I find myself thinking of how much brighter the other students are than I am. (T)
6. I freeze up on things like intelligence tests and final exams. (T)
7. If I were to take an intelligence test I would worry a great deal before taking it. (T)
8. During course examinations, I find myself thinking of things unrelated to the actual course material. (T)
9. During a course examination, I frequently get so nervous that I forget facts I really know. (T)
10. If I knew I was going to take an intelligence test, I would feel confident and relaxed beforehand. (F)
11. I usually get depressed after taking a test. (T)
12. I have an uneasy, upset feeling before taking a final examination. (T)
13. When taking a test, my emotional feelings do not interfere with my performance. (F)
14. Getting a good grade on one test doesn't seem to increase my confidence on the second. (T)
15. After taking a test I always feel I could have done better than I actually did. (T)
16. I sometimes feel my heart beating very fast during important tests. (T)

The remaining subjects received these Nonthreat instructions:

> You know, everyone talks about college students, and what they are like, but few people have actually tried to find out anything from the students themselves. Of course, much time is spent on opinion surveys, and much information is gathered concerning attitudes and opinions of people on various subjects and issues. This kind of information is *not* what I want to obtain here. I am interested in getting an idea of how students think and feel about *themselves*. I'd like to know what you think you are like as a person. I believe the best way to accomplish this is simply to ask you to talk about and describe yourself.
>
> In this experiment, your instructions are to describe yourself, what *you* think about yourself and how you feel about yourself. Naturally I'll say nothing to interrupt you since I'm interested in

TABLE 2

MEAN DIFFERENCES IN PERCENTAGE OF POSITIVE SELF-REFERENCES BETWEEN LAST THIRTY MINUTES OF SESSION AND FIRST TEN MINUTES OF SESSION
(Four subjects per group)

Group	*M*
M-H-T-PSR	.90
F-H-T-PSR	−.30
M-H-T-NSR	−13.03
F-H-T-NSR	1.64
M-H-T-NR	−.91
F-H-T-NR	−5.50
M-H-N-PSR	2.77
F-H-N-PSR	−2.18
M-H-N-NSR	7.39
F-H-N-NSR	−.73
M-H-N-NR	−12.81
F-H-N-NR	−5.72
M-L-T-PSR	1.14
F-L-T-PSR	−2.33
M-L-T-NSR	−9.76
F-L-T-NSR	−5.99
M-L-T-NR	−4.14
F-L-T-NR	−10.26
M-L-N-PSR	8.58
F-L-N-PSR	−2.01
M-L-N-NSR	−7.83
F-L-N-NSR	−.45
M-L-N-NR	3.62
F-L-N-NR	−14.30

Note.—First letter of group label refers to male or female subjects; second refers to High or Low TAS groups; third letter refers to Threat or Nonthreat condition; and the fourth designation to Reinforcement condition, PSR, NSR, and NR, Nonreinforcement.

listening to *you* talk. I'm going to record your description here on this tape recorder—these tapes are identified by code numbers and are used *only* for this project, so no one will know who gave which self-description. You'll have a little over half an hour for this: I'll tell you when to stop. Do you have any questions?

Please hold this microphone. Please try not to tap it on the table, as it is very sensitive and will pick up unwanted noise. OK? Let's begin.

Following administration of the instructions, the 40-minute experimental interview was begun. The first 10 minutes of the interview was a baseline period employed to establish the subject's operant level for emission of positive and negative self-reference statements. No reinforcement or comment was given by the experimenter during this time. During the remainder of the interview and depending on the group to which each subject was assigned, subjects were reinforced either for positive self-references, negative self-references, or not reinforced at all.

Self-references were defined in a manner similar to that employed by Rogers (1960), i.e., descriptive statements about the individual, which tell something about him, or refer to affects which he experiences. As was done by Rogers, four classes of response were scored in analyzing each subject's protocol. These were:

1. Positive Self-References (PSR). Self-references reflecting favorably on the individual (e.g., "I am a well-adjusted person," "I seldom worry," "I like being cooperative").

2. Negative Self-References (NSR). Self-references reflecting unfavorably on the individual (e.g., "I have a problem," "In high school I had great fears about meeting new people," "I am cowardly").

3. Ambiguous Self-References (ASR). Self-references not classifiable as either favorable or unfavorable (e.g., "My major is chemistry," "We talked a lot together," "I am an average student").

4. Other References (OR). Statements with no self-references (e.g., "Mr. Jones is a pleasant person," "He dislikes parties," "They work hard").

The experimenter verbally reinforced positive or negative statements with "um-hmm," "I see," "Yeah," and "uh-huh." In order to be scored as a self-reference it was required that a statement begin with or include I, We, Me, or My, unless the statement was directly related to an immediately preceding statement. No reinforcements were given for statements of a factual nature unless they were associated with affect. In general, the statement classification procedure of Rogers (1960) was employed. Rogers reported an interrater reliability correlation of .88. Pilot studies related to the present research approached this level of reliability.

Experimental design. There were four variables in this research: Test Anxiety Scores—high and low; Experimental Instructions—threat and nonthreat; Sex—male and female subjects; and Reinforcement—subjects reinforced either for positive or negative self-references, or not at all.

In terms of the analysis of variance design used there were four subjects under each of the 24 combinations of these four variables.

RESULTS

Four analyses of operant level verbalizations were performed, one each for the percentage of PSR, NSR, ASR, and OR during the first 10 minutes. In each analysis, the four variables manipulated and their interactions were evaluated. None of the *F*s in the PSR, ASR, and OR analyses attained significance. In the case of NSR, there was a significant *F* for the test anxiety variable ($F = 6.39$, $df = 1/88$, $p < .025$). High TAS subjects had a higher mean percentage of NSR (21.15) than did low TAS subjects (14.98).

The measure employed to determine what effect reinforcement had on the subject's verbalization was the difference between the percentage of statements of a given class during

the last 30 minutes of the session and the percentage of statements of that class during the first 10 minutes. In the case of PSR, a 2 × 2 × 2 × 3 analysis of variance yielded two significant *F*s. One of these was for the variable of Reinforcement ($p < .05$) and the other was the highest order interaction involving all four variables ($p < .05$). Table 2 presents the means of the 24 group combinations of these variables, and Table 3 presents the analysis of variance summary. The means which were compared in the Reinforcement *F* revealed that the subjects reinforced for PSR showed the only positive differences in percentages of PSR between the reinforced and operant periods of the experimental session ($\bar{X} = 1.82\%$). The subjects reinforced for negative sentences showed a mean decrease of PSR of 3.60%; the nonreinforced showed a mean decrease of 6.25%. Further analysis showed, that, of the three reinforcement groups, only the difference between the subjects reinforced for positive statements and those not reinforced at all was significant ($p < .01$). The difference between subjects reinforced for positive and negative statements tended towards significance ($p < .10$). This would indicate that reinforcement of PSR did, but that reinforcement of NSR and nonreinforcement did not, prevent a decrease in PSR during the session. When the high and low TAS subjects reinforced for PSR were compared under the Threat and Nonthreat conditions, no significant differences emerged.

The reinforcement variable was significant, also, in the case of NSR ($p < .01$). The only increase in NSR during the reinforced period as compared with the operant period occurred in the case of subjects reinforced for NSR (2.63%); both groups of subjects reinforced for PSR and those not reinforced at all showed decrements in NSR (−6.27% and −3.18%, respectively). The difference in NSR change between the group reinforced for NSR and the one reinforced for PSR was significant at the .01 level, the difference between the group reinforced for NSR and the one not reinforced was significant at the .05 level. The difference between the group rein-

TABLE 3

SUMMARY OF ANALYSIS OF VARIANCE OF DIFFERENCES BETWEEN PERCENTAGE OF POSITIVE SELF-REFERENCES IN LAST THIRTY MINUTES AND FIRST TEN MINUTES OF EXPERIMENTAL SESSION

Source	*df*	*MS*	*F*
TAS	1	38.72	
Sex (S)	1	96.46	
Instructional condition (C)	1	103.15	1.03
Reinforcement (R)	2	408.38	4.07*
TAS × S	1	148.33	1.48
TAS × C	1	28.35	
TAS × R	2	78.30	
S × C	1	151.53	1.51
S × R	2	248.15	2.47
C × R	2	144.02	1.44
TAS × S × C	1	.04	
TAS × S × R	2	122.78	1.22
TAS × C × R	2	91.49	
S × C × R	2	45.62	
TAS × S × C × R	2	315.19	3.14*
Error	72	100.31	

* Significant at .05 level.

TABLE 4

MEAN DIFFERENCES IN PERCENTAGE OF NEGATIVE SELF-REFERENCES BETWEEN LAST THIRTY MINUTES OF SESSION AND FIRST TEN MINUTES OF SESSION
(Four subjects per group)

Group	*M*
M-H-T-PSR	−10.55
F-H-T-PSR	−5.87
M-H-T-NSR	14.47
F-H-T-NSR	3.81
M-H-T-NR	−10.63
F-H-T-NR	5.93
M-H-N-PSR	−14.27
F-H-N-PSR	−10.05
M-H-N-NSR	.90
F-H-N-NSR	−5.34
M-H-N-NR	−9.27
F-H-N-NR	−.33
M-L-T-PSR	−5.76
F-L-T-PSR	−3.52
M-L-T-NSR	−.54
F-L-T-NSR	−1.40
M-L-T-NR	−1.57
F-L-T-NR	−4.21
M-L-N-PSR	−1.14
F-L-N-PSR	1.04
M-L-N-NSR	3.13
F-L-N-NSR	5.99
M-L-N-NR	−7.18
F-L-N-NR	1.78

Note.—First letter of group label refers to male or female subjects; second refers to High or Low TAS groups; third letter refers to Threat or Nonthreat condition; and the fourth designation to Reinforcement condition, PSR, NSR, and NR Nonreinforcement.

TABLE 5

SUMMARY OF ANALYSIS OF VARIANCE OF DIFFERENCES BETWEEN PERCENTAGE OF NEGATIVE SELF-REFERENCE IN LAST THIRTY MINUTES AND FIRST TEN MINUTES OF EXPERIMENTAL SESSION

Source	*df*	*MS*	*F*
TAS	1	129.07	
Sex (S)	1	152.23	
Instructional condition (C)	1	37.01	
Reinforcement (R)	2	652.36	6.81**
TAS × S	1	3.80	
TAS × C	1	525.52	5.49*
TAS × R	2	258.65	2.70
S × C	1	22.42	
S × R	2	276.82	2.89
C × R	2	20.96	
TAS × S × C	2	59.49	
TAS × S × R	2	120.44	1.26
TAS × C × R	2	38.65	
S × C × R	2	9.43	
TAS × S × C × R	2	126.64	1.32
Error	72	95.76	

* Significant at .05 level.
** Significant at .01 level.

forced for PSR and the nonreinforced group was not significant.

Another effect which was significant ($p < .05$) was that for the interaction between Test Anxiety Scale scores and the Threat-Nonthreat conditions. Whereas the difference between high TAS Threat and Nonthreat conditions (means of −.47 and −6.39, respectively) was significant ($p < .05$), a similar comparison for low TAS group yielded nonsignificant results. Under the Nonthreat but not the Threat condition, the high and low TAS groups were significantly different in mean percentage changes (.60 and −6.39, respectively) at the .05 level.

In order to test the hypothesis that high TAS would react to the Threat condition with heightened responsiveness to reinforcement, comparisons were made among just the groups reinforced for NSRs. It was found that high TAS subjects under the Threat condition showed greater percentage increases in NSR as a function of reinforcement than did the high TAS Nonthreat subjects (means of 9.14 and −3.12, $p < .05$), and, similarly, greater increases in NSR than for all other subjects reinforced for NSR combined ($p < .05$). Table 4 presents the means for percentage change for the 24 groups studied in the NSR comparisons; Table 5 contains the summary of the $2 \times 2 \times 2 \times 3$ analysis of variance.

In the case of the ASR analysis of variance, there were no significant *F*s. In the case of the category of OR, i.e., references not pertaining to the individual, the effect for Reinforcement was significant ($F = 4.04$, $df = 2/72$, $p < .05$). The mean percentage increase in OR for subjects reinforced for PSR was 13.97; in the case of subjects reinforced for NSR the increase was 12.16. Both of these groups had significantly greater mean percentage increases in OR than did the nonreinforced groups (mean of 4.20) (significant at the .01 and .05 levels, respectively).

An analysis was also carried out of the difference in total number of responses between the first 10 minutes and the mean of the last 30 minutes of the experimental session. In this analysis the *F* for Sex was significant (p

TABLE 6

MEAN DIFFERENCES IN TOTAL NUMBER OF RESPONSES BETWEEN LAST THIRTY MINUTES OF SESSION AND FIRST TEN MINUTES OF SESSION
(Four subjects per group)

Group	*M*
M-H-T-PSR	17.75
F-H-T-PSR	13.75
M-H-T-NSR	16.42
F-H-T-NSR	8.00
M-H-T-NR	6.42
F-H-T-NR	4.00
M-H-N-PSR	14.25
F-H-N-PSR	18.00
M-H-N-NSR	14.83
F-H-N-NSR	7.00
M-H-N-NR	19.00
F-H-N-NR	−3.42
M-L-T-PSR	18.67
F-L-T-PSR	14.58
M-L-T-NSR	14.92
F-L-T-NSR	3.50
M-L-T-NR	6.84
F-L-T-NR	6.75
M-L-N-PSR	19.83
F-L-N-PSR	3.75
M-L-N-NSR	5.75
F-L-N-NSR	13.92
M-L-N-NR	16.75
F-L-N-NR	−1.08

Note.—First letter of group label refers to male or female subjects; second refers to High or Low TAS groups; third letter refers to Threat or Nonthreat condition; and the fourth designation to Reinforcement condition, PSR, NSR, and NR Nonreinforcement.

$< .01$) with males giving more responses than did females. Another main effect, the F for Reinforcement, was significant at the .05 level. While all three groups, i.e., those reinforced for PSR, for NSR, and those not reinforced, showed mean increases in total number of responses, the subjects reinforced for PSR had a significantly greater increase than did the nonreinforced subjects (means of 15.07 and 6.91, respectively, $p < .01$). There was a tendency in the direction of a greater increase in total number of responses for the PSR groups than for the NSR group (mean of 10.54). A second significant F in this analysis occurred in the case of the interaction among the variables of Sex, Reinforcement, and Threat-Nonthreat ($p < .05$). Further analyses revealed that this interaction was attributable to: the interaction of Sex and Threat-Nonthreat for nonreinforced subjects ($p < .05$), the interaction of Sex and Reinforcement for subjects under the Nonthreat condition ($p < .01$), and the interaction of Threat-Nonthreat and Reinforcement for male subjects ($p < .05$). Table 6 contains the mean changes in total number of responses for the 24 groups studied. Table 7 contains a summary of the analysis of variance.

TABLE 7

SUMMARY OF ANALYSIS OF VARIANCE OF DIFFERENCES BETWEEN TOTAL NUMBER OF RESPONSES IN LAST THIRTY MINUTES AND FIRST TEN MINUTES OF EXPERIMENTAL SESSION

Source	*df*	*MS*	*F*
TAS	1	23.33	
Sex (S)	1	1138.92	10.11**
Instructional condition (C)	1	1.49	
Reinforcement (R)	2	535.60	4.76
TAS × S	1	.00	
TAS × C	1	15.57	
TAS × R	2	19.61	
S × C	1	79.43	
S × R	2	86.72	
C × R	2	32.70	
TAS × S × C	1	.38	
TAS × S × R	2	153.35	1.36
TAS × C × R	2	50.80	
S × C × R	2	422.85	3.75*
TAS × S × C × R	2	188.82	1.68
Error	72	112.64	

* Significant at .05 level.
** Significant at .01 level.

In an effort to determine the subject's awareness of the reinforcement variable in the interview, each subject was asked the following questions:

1. What do you think the purpose of this experiment might have been?
2. What did you talk about most during the experiment?
3. Did you notice anything I said or did during the experiment? What connection did it have with the experiment?
4. Do you believe that I have influenced your statements in any way? How?
5. Did you notice any changes in the kinds, or types of statements you were making?
6. Do you have any other comments to make concerning this experiment?

On the basis of the subject's answers to these questions a rating of awareness was made for each subject. The rating scale used ranged from 0 to 6. The reliability of this rating scale was found to be .861. The mean awareness rating was .53. On the basis of these ratings and inspection of the subject's answers to the questions it seemed clear that none of the subjects could verbalize the contingency between reinforcement and their verbalizations. However, reinforced subjects had a significantly ($p < .05$) higher mean awareness rating (1.03) than did the nonreinforced subjects (.05). It would seem, then, that subjects displayed some awareness of the reinforcement-verbalization contingency, but that the level of this awareness was quite low.

DISCUSSION

The present findings for the operant period provide strong support for the hypothesis that HA subjects tend to describe themselves in more negative, unflattering terms than do LA subjects. This result is consistent with other evidence in the anxiety literature (Sarason, 1960) which suggests that HA individuals display lower self-esteem, and are more self-debasing in their verbalization than are individuals with lower anxiety scores.

A second hypothesis tested in this experiment involved the effect of the Threat-Nonthreat instructions on the subject's usage of negative self-references during the postoperant period. It had been assumed that high, as compared with low, TAS scorers would show a relatively stronger tendency to emit negative statements, this effect would be height-

ened by a personally threatening condition, and they would show a relatively stronger tendency to be influenced by the experimenter's reinforcements. If these assumptions were correct, it would then be consistent to predict a stronger response of high TAS subjects to reinforcement of NSRs under the Threat as opposed to the Nonthreat condition. No such difference would be expected of low TAS subjects under the two instructional conditions. When comparisons were made for the subjects reinforced for NSR, the finding bore out these predictions. High TAS subjects showed a significantly greater increase in percentage of NSRs under the Threat than under the Nonthreat condition. There was no significant difference between low TAS subjects under the Threat and Nonthreat condition. Looked at in one way, these findings lead one to wonder why HA individuals, told that their verbalizations could be used to assess maladjustment, not only persist in, but significantly increase their references to negative personal characteristics and self-criticism. It is interesting that high TAS subjects did not show a greater increase in positive self-reference when reinforced for PSR under the Threat condition. Studies of the personality characteristics of individuals differing in anxiety may shed light on the meaning of these findings.

Earlier, it was stated that the present research would relate to two broad issues. One was that of differences in verbalization between individuals differing in anxiety; the other, the study of the effects of reinforcement in an interview-like situation. With respect to the first issue, this experiment indicates that successful predictions can be made concerning differences in verbal behavior between high and low TAS subjects under specifiable conditions. This, however, cannot be taken as evidence that all of the results of the experiment involving anxiety immediately follow from the rationale on the basis of which the experiment was designed. For example, it might have been expected that in addition to the significant difference between high and low TAS groups there would also be a significant interaction between TAS scores and instructions during the operant period. Such a significant interaction did not emerge.

In relating this experiment to previous research note must be taken of a possible inconsistency between the present findings and those obtained by Rogers (1960) for a similar task and involving essentially the same system for categorizing verbal responses. Whereas these findings suggest a relationship between anxiety and change in verbal behavior, Rogers failed to obtain a significant relationship between Taylor Manifest Anxiety scale scores and changes in verbal responses under reinforcement conditions similar to those in the present experiment. While procedural and situational factors might account for this difference in results, perhaps the most likely explanation is the difference in the anxiety measures used in the two studies. The Test Anxiety Scale was selected for the present research because it was felt that it might be more sensitive than other possible anxiety measures to the evaluational aspect of the instructions used. Available evidence indicates that the problem of deciding upon the most desirable instrument for predicting anxiety in a particular situation is one requiring further study (Sarason, 1959).

With respect to the question of the manipulation of verbal behavior, the present results strongly suggest the importance of reinforcement as a determiner of usage of response classes. They provide verification for Rogers' (1960) finding that reinforcement of positive or negative response classes significantly influences emission of the response class reinforced. The analysis of variance performed for non-self-responses (OR) indicates that, whether for positive or negative self-references, reinforcement influences not only self-references but non-self-references as well. Nonreinforcement, on the other hand, did not significantly relate to either self- or non-self-references. This might suggest that the reinforcements used had something of a general effect on the subjects verbalizations. Additional support for this inference is provided in the finding concerning changes in total number of responses as a function of reinforcement. Increasingly, studies of verbal behavior make it clear that the response of a

listener to a speaker can exert considerable influence over the speaker's future verbalizations. The time would appear to be ripe to proceed to a systematic analysis of individual and situational differences, and of experimental conditions which determine the extent of this influence.

SUMMARY

The effects of reinforcement, sex, anxiety, and experimental instructions were studied in an experiment in which subjects were asked to describe themselves in a free verbalization situation. The measure of anxiety was the Test Anxiety Scale (TAS), and the experimental instructions were viewed as differing in their personally threatening characteristics. There were three reinforcement conditions with subjects either reinforced for PSR, or reinforced for NSR, or NR.

During an operant 10-minute period, high TAS subjects emitted higher percentages of NSR than did low TAS subjects. A comparison of a reinforced 30-minute period with the operant period revealed that reinforcement for either NSR or PSR significantly influenced the subject's usage of these response classes. For negative, but not for positive, self-references, TAS scores interacted significantly with experimental instructions. This interaction seemed to be due in large part to a greater resistance to extinction of NSR for high TAS subjects under the personally threatening condition. High TAS subjects also were found to respond more positively to reinforcement of NSR under this condition than did low TAS subjects.

The results indicate the importance of reinforcement as a determinant of verbalizations in an interview-like situation. They also demonstrate the necessity of considering individual difference variables (e.g., anxiety) and situational variables (e.g., experimental instructions) in studying the effects of reinforcement on verbal behavior.

REFERENCES

KRASNER, L. Studies of the conditioning of verbal behavior. *Psychol. Bull.*, 1958, **55**, 148–170.

ROGERS, J. M. Operant conditioning in a quasi-therapy setting. *J. abnorm. soc. Psychol.*, 1960, **60**, 247–252.

SARASON, I. G. Relationships of measures of anxiety and experimental instructions to word association test performance. *J. abnorm. soc. Psychol.*, 1959, **59**, 37–42.

SARASON, I. G. Empirical findings and theoretical problems in the use of anxiety scales. *Psychol. Bull.*, 1960, **57**, 403–415.

(Received August 2, 1961)

Journal of Abnormal and Social Psychology
1962, Vol. 65, No. 5, 308-318

LEADER ATTITUDES, GROUP CLIMATE, AND GROUP CREATIVITY [1]

FRED E. FIEDLER

University of Illinois

The rapid expansion of knowledge within the past few decades has brought with it the need for increased specialization. One consequence of this development has been the greater reliance on teams of experts in the solution of complex problems. This trend is apparent in government, in industry, and in other fields, such as law and architecture. It is especially pronounced in research and development efforts where the effectiveness of a team may have far-reaching effects. Yet, despite the obvious importance of this problem, only a few workers have dealt with it (e.g., Bush & Hattery, 1956; Parnes & Meadow, 1959; Pepinsky, 1959; Taylor, 1958) and very little is known about the determinants which contribute to the productivity of such teams. The present report summarizes four investigations which deal with the effect of the leader's interpersonal attitudes on his group's creativity.

Group creativity is here defined as a group interaction leading to an original solution of a problem or to a more effective recombining of given elements into new forms. This definition does not require the group itself to think or create, although it does assume that group members may stimulate one another, reject poor solutions, and propose improvements.

The main concern of this paper is the prediction of group creativity by certain leader attitudes which are here measured by means of interpersonal perception scales. These scales ask the leader to think of all the people with whom he has ever worked and then to describe the person whom he considers his best co-worker and the person with whom he found it most difficult to work.

The descriptions are made on simple eight-point bipolar adjective scales usually containing 20 scale items,[2] such as:

Very contented :—8—:—7—:—6—:—5—:—4—:—3—:—2—:—1—:Not contented
Very stubborn :—1—:—2—:—3—:—4—:—5—:—6—:—7—:—8—:Not stubborn
Not enterprising:—1—:—2—:—3—:—4—:—5—:—6—:—7—:—8—:Very enterprising

The main predictor score in this study was the Esteem for the Least Preferred Co-worker (LPC). This score is defined as the favorableness with which a subject described the *poorest* co-worker he ever had. It is obtained by assigning the maximum score (i.e., 8) to the most favorable pole and a score of 1 to the least favorable pole of each scale of the LPC description, and summing the scores over all 20 scale items. LPC is a component of the Assumed Similarity between Opposites (ASo) score which is obtained by comparing the profile similarity between the subject's descriptions of his most and least preferred co-workers. This com-

[1] This paper was based on research conducted under United States Public Health Service Grant M-1774 and Contract Nonr-1834(36) with the Office of Naval Research. The writer is especially indebted to D. E. Dulany, J. E. McGrath, I. D. Steiner, H. C. Triandis, L. R. Tucker, and W. A. T. Meuwese for their critical comments, as well as to an anonymous, but unusually conscientious and capable journal reviewer.

[2] The remaining items in this scale were: Not clear thinking-Very clear thinking, Not tense-Very tense, Very abrupt-Not abrupt, Not hardworking-Very hardworking, Not studious-Very studious, Not sympathetic-Very sympathetic, Not patient-Very patient, Not reliable-Very reliable, Not dejected-Very dejected, Very silly-Not silly, Not enthusiastic-Very enthusiastic, Not self-confident-Very self-confident, Not agreeable-Very agreeable, Not productive-Very productive, Not easily discouraged-Very easily discouraged, Not adventurous-Very adventurous, Not lonely-Very lonely, Very sociable-Not sociable, Not dissatisfied-Very dissatisfied, Not ambitious-Very ambitious, Not uncertain of himself-Very uncertain of himself. Other sets of similar adjectives appear to be equally useful.

parison is generally made by means of the *D* statistic (Cronbach & Gleser, 1953). A high Assumed Similarity score (small *D* score) indicates that the subject perceives his co-workers as very similar, while low ASo indicates that the subject perceives his most and least preferred co-workers as very different. The ASo and LPC scores typically correlate about .70 (e.g., Fiedler, Meuwese, & Oonk, 1961).

As this paper will show, the high LPC leader, who perceives even his least preferred co-worker relatively favorably, tends to be more effective in groups which he, at least, experiences as relaxed and pleasant. The low LPC leader, on the other hand, is able to operate more effectively in groups which are seemingly more tense and stressful.

THE DUTCH STUDY:

EFFECTS OF GROUP COMPOSITION AND LEADERSHIP ON PREDICTING GROUP CREATIVITY

The first investigation in the series (Fiedler et al., 1961) explored three different factors which were hypothesized to influence group creativity: the socioreligious background of team members, the difference between groups with appointed and emergent leaders, and the leader's interpersonal perceptions.

Our interest in the first of these variables was aroused by the particular social conditions in Holland, a country with a long history of conflict between the Protestant North and the Catholic South. This cleavage has existed to a greater or lesser extent ever since the Reformation and the 80 years war of independence from Spain, and is especially pronounced between Calvinists and Catholics. Even in cities with a mixed religious population, the limited social and business contacts between members of these different religous groups tend to be formal and somewhat strained. On the other hand, political bodies and governmental commissions are customarily selected so as to represent the viewpoints of various religious groups. Thus, the degree to which the members of these heterogeneous groups can fruitfully collaborate in policy making and decision making has a vital influence on the social and economical well-being of the country.

The second question concerned the group's creativity under appointed or emergent leadership. Several previous investigations have failed to find systematic differences in group performance between groups having formal and informal leaders (Borg, 1957; Carter, Haythorn, Shriver, & Lanzetta, 1951), yet these factors might well interact with others in influencing the creativity of groups.

Third, this study explored the effect of the leader's interpersonal perceptions on group creativity. In previous research, the leader's ASo was highly correlated with group productivity (Fiedler, 1958). However, this earlier work dealt with groups in which the leader directed and coordinated the work of others, e.g., in winning a basketball game, surveying land, producing steel, or in the profitable management of a company. The present investigation, using LPC as well as ASo scores, extended this research to group tasks requiring creativity.

Procedure

The study utilized 32 Calvinist and 32 Catholic first and second year male university students. Each subject completed a set of interpersonal perception scales describing his most and least preferred co-workers. A total of 32 groups, each consisting of four men were formed. Sixteen of the groups consisted of two Catholics and two Calvinists (heterogeneous groups). Seven[3] groups contained four Catholics and 7 groups contained four Calvinists. Each subject participated once in a homogeneous and once in a heterogeneous group. In half of the groups within each category, the experimenters appointed a chairman ("formal" groups), while the other groups were operated under emergent leadership conditions ("informal" groups).

The groups were given a TAT card (Card 11 for heterogeneous and Card 19 for homogeneous groups) and told to devise three stories within 20 minutes. These stories were to be as original and as different as possible. The TAT stories were rated by two independent judges using a scoring manual. Interrater reliabilities were .82 and .89. At the termination of the task subjects completed sociometric questionnaires indicating the informal leaders of each group, i.e., group members who "had most influence on the opinions of others," with whom the subject would most like to work again on a similar task, and whom the subject would choose as a leader on a similar task.

Results

Homogeneous and heterogeneous groups did not differ significantly on the TAT task,

[3] Because of a clerical error, two presumably homogeneous groups contained one man of the other faith and had to be deleted from the sample.

and no significant differences were found between groups under appointed and emergent leaders.

To determine the influence of the leader on group creativity, his ASo and LPC scores were correlated with the group performance. In formally organized groups the chairman's score was utilized while the sociometrically chosen leader's score was used in the informal teams. The rank-order correlations of the leader's ASo or LPC scores with group creativity scores were not significant. However, when the groups were subdivided into homogeneous and heterogeneous groups by leadership type, the following results were obtained. In homogeneous groups with formal leaders, the formal leader's LPC score correlated positively with group performance. In all heterogeneous and informal groups, these correlations were negative (see Table 1).[4] The ASo scores yielded correlations which were in the same direction but considerably weaker. Further analyses were, therefore, based primarily on the esteem for Least Preferred Co-worker (LPC) scores.

These results suggested a complex interaction between LPC, group creativity, and the group's organization. In attempting to interpret these findings we assumed that formal homogeneous groups would tend to be more pleasant and relaxed than the others since homogeneous groups, containing men from the same religion, would make the members less defensive and guarded in their interactions. Likewise, groups having an appointed leader would operate under less stress since the appointment of a leader would make it unnecessary to compete for leadership status. Presumably, then, the formal homogeneous groups would be least stressful and unpleasant, while all other groups would be working under varying degrees of greater stress. This interpretation was supported by the finding that subjects indicated the presence of a "destructively critical group member" in only 14% of formal homogeneous groups but in 50, 67, and 87% of the other groups. This led to the hypothesis that the leader's LPC score correlates positively with group creativity in pleasant groups, while LPC correlates negatively with creativity in groups operating under more socially strained and unpleasant conditions.

AN INVESTIGATION INVOLVING HYPNOTIC MANIPULATION OF LEADER ATTITUDES

Procedure

The second investigation in this series (Fiedler, London, & Nemo, 1961) was in part designed to test the hypothesis which emerged from the Dutch study. The experiment utilized basically identical interpersonal perception tests and criterion measures, but included, in addition, a scale specifically designed to indicate the perceived pleasantness or stressfulness of the group interaction. This Group Atmosphere (GA) scale utilized 10 highly intercorrelated bipolar adjective items, and followed the form of the interpersonal perception tests, such as:

Friendly :—8—:—7—:—6—:—5—:—4—:—3—:—2—:—1—:Unfriendly

Rejecting :—1—:—2—:—3—:—4—:—5—:—6—:—7—:—8—:Accepting

Unhelpful:—1—:—2—:—3—:—4—:—5—:—6—:—7—:—8—:Helpful[5]

The subjects were 24 women, 12 of whom had been classified as highly susceptible to hypnosis, or "tranceable," while the other 12 subjects had scored as very unsusceptible, or "untranceable," on the Stanford Hypnotizability scale (Weitzenhoffer & Hilgard, 1959). Nevertheless, every one of the 24 subjects was assured that she was hypnotizable, and was subjected to identical experimental treatment.

Each of the women received three hypnotic suggestions in the course of the study. These asked her: to perceive her most and least preferred co-workers as being very similar—the High ASo condition; to see them as very dissimilar—the Low ASo con-

[4] It will be noted that the heterogeneous informal groups, which would presumably be most stressful, yielded the lowest correlation. This may be due to the smaller variance in creativity scores and the consequent restriction in range found in this set of groups.

[5] The remaining items were: Unenthusiastic-Enthusiastic, Lots of fun-Serious, Nonproductive-Productive, Distant-Close, Cold-Warm, Cooperative-Uncooperative, Supportive-Hostile, Boring-Interesting, Successful-Unsuccessful.

dition; and, to see them as she would normally think of them—the Norm condition.

After each hypnotic induction, the subject described the most and least preferred co-workers she had ever had. She was then sent into a room where she met two other women who were introduced to her as fellow subjects, but who were in fact confederates of the experimenters. The group was asked to choose a chairman, which always eventuated in the subject's "election." The subject and confederates were then given one TAT card per session on the basis of which they were to devise three different original stories, within a 15-minute period. Cards 11, 19, and 17GF in counterbalanced sequence were utilized, i.e., pictures of a dragon in a ravine, an ice-covered house, and a girl on a bridge (Murray, 1938).

At the end of each session, the subject and confederates described the group atmosphere, and the subject returned to her hypnotist for the next hypnotic induction. The sequence of hypnotic instructions as well as presentation of cards was counterbalanced. Neither the administrator of the experimental sessions nor the confederates knew, or were able to guess, the subject's tranceability or the hypnotic instructions she had received. All sessions were recorded and the tapes were analyzed by Bales' (1950) Interaction Process Analysis. Stories were rated for creativity by three independent judges using a scoring system essentially identical to the one used in the Dutch study. The inter-rater agreement of creativity scores for three raters was .94.

TABLE 1

CORRELATIONS BETWEEN LEADERS' LPC AND PRODUCTIVITY SCORES IN THE DUTCH CREATIVITY STUDY

Group structure	Group composition	
	Homogeneous	Heterogeneous
Formal	No competition for leadership status Little social strain; few problems in communication $\bar{X}$[a] 105.28 SD[a] 37.23 N 7 $rho_{LPC\ Creativity}$ +.75*	No competition for leadership status Much social strain; impeded communication $\bar{X}$ 116.62 SD 34.37 N 8 $rho_{LPC\ Creativity}$ −.72**
Informal	Competition for leadership status Little social strain; few problems in communication. $\bar{X}$ 87.50 SD 16.64 N 6 $rho_{LPC\ Creativity}$ −.67	Competition for leadership status Much social strain; impeded communication $\bar{X}$ 105.75 SD 10.30 N 8 $rho_{LPC\ Creativity}$ −.21

[a] Means and standard deviations refer to creativity scores.
* $p < .10$.
** $p < .05$.

TABLE 2

CORRELATIONS BETWEEN LEADER'S LPC AND GROUP CREATIVITY IN THE HYPNOSIS STUDY ($N = 8$)

ASo condition	Top third	Middle third	Lower third
	Leader's GA scores		
High	.22	.36	−.40
Low	.24	−.24	−.71**
Norm[a]	.64**	−.33	−.72***
	Confederates' GA scores		
High	.17	.52	.00
Low	.43	−.22	−.14
Norm[a]	.27	−.28	−.62

[a] The Norm condition is most representative of the subject's usual mode of perceiving co-workers.
** $p < .05$, one-tailed.
*** $p < .025$, one-tailed.

Results

The sessions were divided into those scored by the subject, or by confederates, as being relatively pleasant, neutral, or unpleasant on the group atmosphere scales. (The leader's LPC and GA scores were uncorrelated—$N = 24$, −.08.) As hypothesized, in groups which were seen by the subject as pleasant and relaxed, LPC correlated positively with group creativity. In group sessions which the subject experienced as stressful and unpleasant, LPC and group creativity correlated negatively. The predicted direction of results held for all three experimental conditions and correlations reached the .05 level of significance for three of the six critical comparisons. These relations were somewhat less systematic when the groups were divided according to group atmosphere scores of confederates (see Table 2).

These findings raised a number of important questions. First of all, the criterion of group creativity in both investigations was based on TAT stories which were scored by essentially identical methods. Evidence demonstrating the generalizability of these findings to other criteria was clearly re-

quired. Second, since the group members were confederates, the study failed to answer whether the group members' assessment of the group atmosphere under more natural conditions was more important in determining the optimal relation between LPC and group creativity. Third, the "group members" in this study had been instructed to act in certain ways, e.g., they were to cooperate fully with the subject but to refrain from introducing new ideas or procedures. Hence, it was essential to determine whether "group creativity" in this investigation really reflected the group's performance or merely the originality of the subject. The third study in this series provided an opportunity to test these questions.

THE LEADERSHIP TRAINING STUDY: A CROSS-VALIDATION

Procedure

This third investigation (Fiedler, Bass, & Fiedler, 1961) was conducted as part of a national leadership training conference which brought together experienced church administrators and lay leaders of the American Unitarian Association. The group consisted of 71 women and 37 men, most of whom were between 35 and 50 years of age. The conference devoted five consecutive mornings to leadership training. In the course of this program, subjects were assembled into five-person teams. Each team consisted of a designated chairman, three members, and an "observer" who did not participate in the group discussions but administered questionnaires, timed the session, and directed the postsession critique.

New groups were formed each day so that each person served as a leader on one day and as an observer one day, and no two group members worked together on more than two days. The group composition on the first and second days was completely different. Moreover, the creativity score of the group of which an individual was a member on one day was uncorrelated with the performance of groups in which he served on succeeding days. Thus, the groups can be considered independent cases.

On the first day, before being assigned to any groups, subjects completed interpersonal perception scales from which LPC scores were derived. Immediately after each session, subjects rated the group's atmosphere on the same scale used in the hypnosis study.

Group creativity tasks. All groups worked on the same task on the same day, and a different task was presented on each successive day. Four of these tasks were designed to test the groups' creativity in different situations. The fifth was a human relations problem and is not relevant to this study. The tasks were to be completed within a 35-minute period.

On Monday the groups were to write a 250-word statement justifying a decision by the congregation's minister not to divulge the intent of a local physician (not a church member) to commit a mercy killing.

On Tuesday, the groups had to invent a Sunday school parable for 6–8 year olds which would illustrate the desirability of the doctrine of separation of church and state.

The Wednesday groups were asked to develop an appeal for funds to send a ministerial student through divinity school.

On Thursday the groups planned a 3-minute skit for presentation that evening to illustrate the need for better music in the worship service.

These problems were designed to tap different aspects of group creativity, and they also differed from the TAT tasks which had been used in the preceding studies. The criterion of group performance consisted in each case of the judgments and ratings of all other conference participants. (Subjects did not rate the product of their own group.)

The reliability of the criterion ratings was assessed by randomly dividing the ratings into two subgroups and computing separate rankings. The rank-order correlations between these two sets of ratings were .91, .76, and .90 for Monday, Tuesday, and Thursday, respectively. The criterion score for the fund raising problem involved putting play money into collection boxes; this precluded the computation of a reliability estimate.

Results

As in the previous experiment, positive correlations between the leader's LPC score and group creativity were expected in teams in which the leader experienced the climate as being pleasant. Negative correlations were anticipated in teams where the leader experienced the group climate as unpleasant and stressful. Since no absolute norms were available for interpreting GA scores, the groups were again ranked on the basis of the leader's GA scores, as well as on the group members' GA scores, and divided into an upper, middle, and lower third.

The correlations between the leaders' LPC scores and group creativity under the three GA conditions are listed in Table 3. As can be seen, the consistency in the patterns of these correlations supported the hypothesis on each of the four days. While the samples were very small, and the correlations individually did not reach the .05 level of significance, the results were quite consistent from day to day. All GA scores were rela-

TABLE 3

CORRELATIONS BETWEEN LEADER'S LPC SCORES AND GROUP CREATIVITY UNDER HIGH AND LOW LEADER GA SCORES IN THE LEADERSHIP TRAINING STUDY

	Leader's GA scores		
	Top third	Middle third	Lower third
Monday (Justification)	.28	.10	.03
Tuesday (Parable)	.89***	.67	−.03
Wednesday (Fund appeal)	.14	.19	−.40
Thursday (Skit)	.37	−.08	−.60

*** $p < .025$.

tively high on Monday, and on that day the correlation between LPC and task performance was .03 even for the least pleasant groups. The three variables, leader LPC, leader GA, and Group Creativity, are statistically independent measures. Taken over all groups, the median intercorrelations for the four days were −.18 for LPC and GA, .06 for GA and Creativity, and .03 for LPC and performance if the groups are not subdivided on the basis of the leaders' GA scores.

Primarily to illustrate that the correlation between the leader's LPC score and Group Creativity is more positive as the leader's GA score increases, the leader's average GA scores within each cell in Table 3 and the corresponding correlation coefficients between LPC and Creativity were correlated. This correlation was .60 ($N = 12$) which is significant in the expected direction at the .025 level of confidence. Hence, the direction and magnitude of the LPC-Group Creativity relationship is related to the leader's GA score. As was the case to some extent in the previous study, the relations between LPC and Group Creativity were not influenced by group members' GA scores. Here the corresponding correlation was −.26.

The major significance of this third study was in the cross-validation and extension of the first two experiments. The last study extended the earlier work by utilizing adult subjects with previous leadership experience, and by testing the hypothesis with four different tasks which closely resemble the problems with which these subjects normally deal.

Since this series of studies was conducted with ad hoc groups, we must ask whether these complex relations also occur in natural groups under real-life conditions. A reanalysis of data from a previous investigation made it possible to assess the status of our major hypothesis on such groups having policy making and decision making functions which require some measure of creativity.

REANALYSIS OF A STUDY OF BOARDS OF DIRECTORS

An earlier study (Godfrey, Fiedler, & Hall, 1959) obtained interpersonal perception and group performance data on boards of directors of 32 very similar farm supply cooperatives for which unusually reliable and valid criteria of organizational effectiveness were available. This investigation showed that the company's success was strongly influenced by the general manager, whose ASo score correlated very highly (i.e., up to −.74) with the company's net income, taken over a 3-year period. In contrast, the relations between the Board Leader's ASo and company performance criteria were relatively

TABLE 4

SOME RELATIONS BETWEEN BOARD PRESIDENTS' LPC SCORES AND COMPANY NET INCOME IN THE FARM SUPPLY COMPANY STUDY[a]

	N	Correlation LPC of board's leader and net income
A Board President and General Manager who is accepted by staff sociometrically endorse each other	5	.67
B[b] President does not endorse General Manager	8	−.40
C General Manager sociometrically endorses President who is informal leader of Board	11	.14
D General Manager does not sociometrically endorse President who is informal leader of Board	7	−.51
E Board sociometrically chooses General Manager	10	.21
F Board does not sociometrically choose General Manager	10	−.60**

[a] The number of overlapping cases between samples is as follows: A and E, 1; A and C, 4; C and E, 1; B and D, 1; B and F, 1; D and F, 2; A and D, 1; A and F, 2; B and C, 4; B and E, 4; D and E, 1.

[b] In no case did Board President and endorsed General Manager sociometrically reject each other.

** $p < .05$.

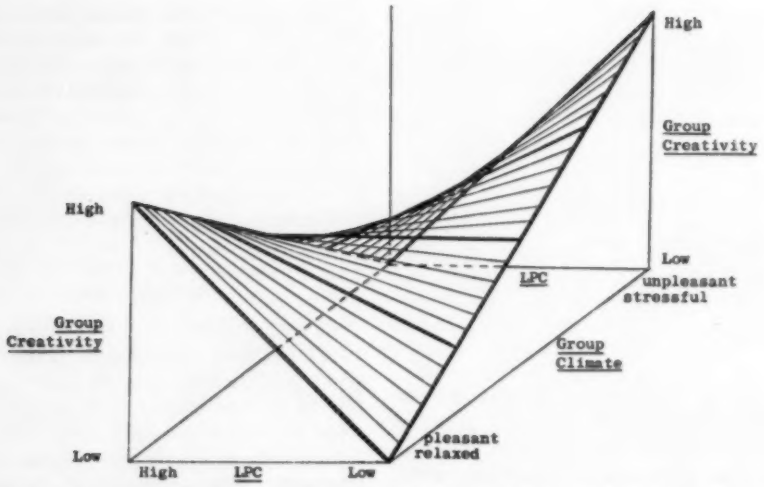

FIG. 1. Correlation surface representing the correlations Leader's LPC and Group Creativity under different conditions of group stress as experienced by the leader.

weak when compared with those obtained for general managers.

For the purpose of the present analysis, we now computed the LPC scores of all leaders of the boards. The companies were divided into those in which the presence of conflictful relations between boards and managers could reasonably be inferred and those in which interpersonal relations appeared to be harmonious. Specifically, we compared companies in which the board leader sociometrically endorsed the general manager, and those in which he did not. Similarly, we compared companies in which the boards sociometrically endorsed the general managers, and those in which they did not. The three criteria of pleasant versus tense board-manager relations are relatively independent. While only one of the correlations reached an acceptable level of significance, all six of the correlations were in the direction supporting the major hypothesis of this paper (Table 4).

DISCUSSION

Our studies yielded consistent relations showing that the leaders of effective groups tended to see least preferred co-workers and group atmosphere either both as favorable or both as unfavorable. Where the leader's perceptions were dissonant the group tended to be less effective. These findings can be cast in the form of a model which incorporates the three main variables of these studies, viz., the leader's LPC and GA scores, and the group's creativity. The resulting correlation surface shows the nature of this interaction (see Figure 1).[6] While this model is far more rigorous than the data warrant, it may serve as a point of departure for future research. Other determinants, such as task dimensions or individual member abilities, will undoubtedly need to be incorporated in it as further research becomes available.

The results of these studies raise a number of questions. The most important of these concern the behavioral correlates of LPC and group performance, the meaning of group atmosphere scores, and the interaction of LPC and GA scores in predicting group creativity.

Behavior associated with LPC and Group Creativity scores. Content analyses were performed in the Dutch and Hypnosis studies in the hope that they might shed light on these relations. Since the analyses had to be based on very small *N*s the results, at best, are

[6] Thanks are due to Raymond W. Frankmann, Jr., for suggesting this formulation.

inconclusive and must be seen primarily as hypotheses for further research.

In the Dutch study we categorized remarks by all members of a group into five categories, viz., procedural remarks ("You start the story.") new ideas; elaboration of ideas; critical comments; and irrelevant and tension releasing remarks. In addition, a count was made of the total number of remarks in each group (Meuwese & Oonk, 1960). To parallel the studies for which GA scores were available, the teams were identified as "relaxed" or "tense." A group was considered "tense" if a high proportion of "critical remarks" had been made by participants during the session, and if one or more members in the group had been named as "destructively critical" on the sociometric questionnaire. The comparison of the 10 most and 10 least "relaxed" groups is presented in Table 5. For purposes of this discussion we shall interpret correlations in the neighborhood of .30 or above.

The first two columns in Table 5 show (*a*) that procedural and irrelevant comments tended to aid the creativity of relaxed groups but hindered tense groups; (*b*) the more task oriented "elaboration of ideas" and higher total activity aided the tense groups but not relaxed groups. Looking at the right-hand columns of this table, we find that the more person oriented, procedural comments (Category 1) and irrelevant remarks (Category 5) were relatively more common in groups of high LPC leaders while the task relevant elaboration of ideas was more common in groups of low LPC leaders. Hence, the more person oriented behaviors characteristic of groups under high LPC leaders helped in relaxed situations; the tense groups fared better with relatively more task oriented behavior.

More detailed content analyses were made of the Norm condition in the Hypnosis study. Leader comments were categorized according to Bales' (1950) Interaction Process Analysis. The first two columns of Table 6 present the correlations between group creativity and the proportion of leader comments in each category. As can be seen, "positive social-emotional" comments (Categories 1 and 2) were conducive to group creativity while tension

TABLE 5

CORRELATIONS OF CONTENT CATEGORIES WITH LEADER LPC AND GROUP CREATIVITY SCORES IN THE DUTCH CREATIVITY STUDY

(Ns = 10)

Category	Group creativity		Leader's LPC scores	
	Relaxed groups	Strained groups	Relaxed groups	Strained groups
1. Procedural remarks	.48	−.44	.54	.40
2. New ideas	−.71	−.43	−.11	−.14
3. Elaboration	−.18	.46	−.28	−.59
4. Critical remarks	.25	.00	−.15	.08
5. Irrelevant comments	.53	−.25	.27	.61
6. Total activity	−.64	.48	.05	−.52

Note.—$p < .10$ at .56; $p < .05$ at .64.

indicating "negative social-emotional" comments (Categories 10, 11, and 12) were deterimental.

The right-hand columns of Table 6 present the correlations between LPC and the Bales (1950) categories. These show that the high LPC leaders tended to react in a tension free manner (Categories 2, 10, and 11) in groups which they perceived as pleasant, but they acted tense, disagreeable, and antagonistic (Categories 10, 11, and 12) in groups they saw as relatively stressful. In contrast, the low LPC leaders reacted *with more tension* in groups they saw as *pleasant* than in groups they rated as unpleasant. In addition, pleasant groups seemed to require more compliant, nondirective leaders (Categories 3 and 4) while the less pleasant, low GA groups seemed to require more positive, directive leadership (Categories 3, 5, and 6). These relations are presented schematically in Table 7.

In brief, the content analyses of the Dutch and Hypnosis studies lead to the following interpretation: (*a*) Group creativity requires tension free leadership behavior of a nondirective kind in pleasant groups, and of a directive, task oriented nature in groups which are relatively more stressful; (*b*) High LPC leaders show tension free and nondirective behavior in relaxed groups while Low LPC leaders show tension free and task oriented behavior in more stressful situations. Thus, High LPC leaders behave in manner which is detrimental to unpleasant groups; Low LPC leaders behave in a manner which is detri-

TABLE 6

CORRELATIONS OF LEADER'S INTERACTIONS[a] (BALES) WITH GROUP CREATIVITY AND LPC SCORES IN THE HYPNOSIS STUDY

(Ns = 8[b])

Category	Creativity		LPC	
	Pleasant	Unpleasant	Pleasant	Unpleasant
1. Shows solidarity, encourages, praises	.10	.70	.17	−.31
2. Shows tension release, jokes, laughs	.13	.31	.53	−.05
3. Agrees, concurs, complies	.52	−.45	.14	.33
4. Gives suggestion, direction	−.55	.00	−.45	−.58
5. Gives opinion, evaluation	.19	−.43	.26	.47
6. Gives orientation, information	−.05	.48	−.26	−.14
7. Asks for orientation, information	.19	.17	.64	−.24
8. Asks for opinion, evaluation	−.17	−.08	−.28	.04
9. Asks for suggestion, direction	.00	.29	−.72	−.35
10. Disagrees, shows rejection	−.43	−.36	−.50	.61
11. Shows tension, withdraws	−.78	−.48	−.64	.41
12. Shows antagonism, defends self	−.28	−.14	.00	.34
Total Interaction	−.12	.28	.02	−.53

[a] Proportion of comments.
[b] Rank-order correlations of .73 and .82 are required to reach the .05 and .02 levels, respectively, for Ns of 8.

mental to pleasant groups. It should be noted, however that none of these situations could be considered extremely stressful or anxiety arousing.

Meaning of GA scores. The leader's GA rating might reflect the degree to which his needs for passive or dominant behavior were satisfied or it might reflect the actual pleasantness of the group. The latter seems somewhat more likely although it is possible that low LPC leaders may have been happier in groups they could dominate and in which they could express their antagonisms and tensions (even though this behavior was detrimental to the task). The data suggest, however, that the leaders' perceptions were in part situationally determined.

First, it will be recalled that social strain was experimentally introduced in the Dutch study by varying group composition and

TABLE 7

EFFECTS OF LEADER'S BEHAVIOR ON GROUP CREATIVITY IN THE HYPNOSIS STUDY UNDER HIGH AND LOW GROUP ATMOSPHERE

Leader's perceived group atmosphere	Beneficial	Leader's LPC[a]	Detrimental	Leader's LPC[a]
Pleasant (High GA_L)	3. Agrees, concurs, complies	—	4. Gives suggestions, directions	Low
			10. Disagrees, rejects	Low
			11. Shows tension	Low
Unpleasant (Low GA_L)	1. Shows solidarity, gives help praises	Low	3. Agrees, concurs, complies	High
			5. Gives opinions, evaluations	High
	2. Shows tension release, jokes, laughs	—	10. Disagrees, rejects	High
	6. Gives orientation, information	—	11. Shows tension, withdraws	High

[a] This column indicates the LPC scores of leaders which tended to be associated with the behavior listed in the cell.

leadership, and it was inferred from sociometric choice patterns in the Farm Supply Company study. As Table 4 showed, the definition of strained board-management relations did not rest exclusively on the board leader's choices. The fact that similar interactions were obtained in both the Dutch and Farm Supply Company studies as in those using leader GA scores suggests, therefore, that the same social factors may underlie the different measures of stress, and that the leader is sensitive to them.

Second, we must ask whether the effective leaders tend to create relaxed or stressful group climates to fit their own leadership style. This, again, seems unlikely. Analyses from a current study by McGrath and Julian (unpublished manuscript) have shown that the leader's GA scores are uncorrelated over different tasks and groups. This finding suggests that the leader does not "make" the group atmosphere pleasant or unpleasant. Obviously, however, an understanding of the part played by group atmosphere or stress requires further research.

Interaction of leader LPC and GA scores in predicting group creativity. Even though most of the individual correlations, using content analyses, were not statistically significant, the patterns explicated the nature of the interaction of the leader's LPC and GA scores in predicting group creativity. Why, however, is a group more creative when the leader's perception of LPC and the group climate are congruous than when they are not? Several explanations are possible, although all of them are highly tentative.

One interpretation would be that the High LPC leader is more concerned with good interpersonal relations and becomes tense and anxious, and hence less effective, in situations involving interpersonal stress. However, Low LPC leaders who are generally more concerned with the task success of their groups become tense and anxious in a pleasant, relaxed interpersonal climate which they may see as a threat to group success.

This explanation accords with the common experience that some people are very effective in emergencies or under pressure where the urgency of the task permits them to relegate concern about good personal relations to secondary importance. These same individuals frequently cannot keep a group together when the sailing is smooth. Other leaders operate well under normal conditions although they cannot deal with an emergency. And we did, in fact, find a tendency of group members in the Hypnosis study to rate the leaders as more self-assured under congruent than noncongruent conditions.

The data, and the interpretations, leave many questions unanswered and it is hoped that current and future research will clarify many of the issues which remain. The data do, however, suggest a resolution of conflicting leadership theories. Various administrators and psychologists have held that close, personal, nondirective leader attitudes are most conducive to group performance (e.g., Gordon, 1955; McGregor, 1960) while others have argued for the effectiveness of role oriented, psychologically distant leader attitudes (e.g., Fiedler, 1958; Katz, Maccoby, Gurin, & Floor, 1950; Spector & Suttell, 1956). The present research suggests that either of these leadership approaches can be successful if it is compatible with the leader's tolerance for relaxed or strained interpersonal relations.

Limitations of the research. The data within this series of studies are quite consistent and, taken as a group, they are certainly significant well beyond the customarily accepted probability levels. The reader should, however, note certain important limitations of the results which affect the generalizability of the findings:

1. These studies were conducted under generally nonstressful conditions. Even "low Group Atmosphere scores" are likely to indicate only unpleasantness or at most moderate stress. The effects of extreme stress on group performance might well change the nature of the relationships which were found here.

2. The studies were conducted with groups in which the leader was in a chairmanship position, i.e., he was a participant in the group process and had rather limited power over his members. It seems not unlikely that supervisory relations may require different attitudes than do participatory leadership roles.

3. The group creativity tasks are not highly intercorrelated, and these tasks cannot be considered an adequate sampling of all possible tasks. Until we can specify the dimensions of group tasks more precisely than is possible at this time, the prediction of performance to new tasks must remain a risky procedure.

SUMMARY AND CONCLUSIONS

Three laboratory studies and one investigation of natural groups are here reported which relate the leader's perception of his least preferred co-worker (LPC) to the performance of his team on tasks requiring creativity.

The research program provided consistent evidence that leaders who perceive their least preferred co-worker favorably tend to be most effective under pleasant and relaxed group conditions, while leaders who perceive LPC unfavorably are more effective under unpleasant stressful group climates. The same general trends emerged from a reanalysis of data obtained on "real-life" groups, viz., boards of directors and management of 32 consumer cooperatives.

Content analyses of specific leader and group member behaviors suggested that group creativity requires tension free leadership of a nondirective kind in pleasant groups, but of a tension free, directive kind in relatively more stressful, unpleasant groups. High LPC leaders show this behavior in groups they perceive to be pleasant while low LPC leaders tend to manifest such behavior under more stressful conditions. The results lead to the construction of a mathematical model which is briefly discussed along with the theoretical implications of the research.

REFERENCES

BALES, R. F. *Interaction process analysis: A method for the study of small groups.* Cambridge: Addison-Wesley, 1950.

BORG, W. R. The behavior of emergent and designated leaders in situational tests. *Sociometry,* 1957, **20,** 95–104.

BUSH, G. P., & HATTERY, L. H. Teamwork and creativity in research. *Admin. sci. Quart.,* 1956, **1,** 361–372.

CARTER, L. F., HAYTHORN, W., SHRIVER, BEATRICE, & LANZETTA, J. The behavior of leaders and other group members. *J. abnorm. soc. Psychol.,* 1951, **46,** 589–595.

CRONBACH, L. J., & GLESER, GOLDENE C. Assessing similarity between profiles. *Psychol. Bull.,* 1953, **50,** 456–473.

FIEDLER, F. E. *Leader attitudes and group effectiveness.* Urbana: Univer. Illinois Press, 1958.

FIEDLER, F. E., BASS, A. R., & FIEDLER, JUDITH M. The leader's perception of co-workers, group climate, and group creativity: A cross-validation. Urbana: University of Illinois, Group Effectiveness Research Laboratory, 1961. (Mimeo)

FIEDLER, F. E., LONDON, P., & NEMO, R. S. Hypnotically induced leader attitudes and group creativity. Urbana: University of Illinois, Group Effectiveness Research Laboratory, 1961. (Mimeo)

FIEDLER, F. E., MEUWESE, W. A. T., & OONK, SOPHIE. Performance of laboratory tasks requiring group creativity. *Acta psychol., Amsterdam,* 1961, **18,** 100–119.

GODFREY, ELEANOR P., FIEDLER, F. E., & HALL, D. M. *Boards, management, and company success.* Danville, Ill.: Interstate Publishers, 1959.

GORDON, T. *Group centered leadership.* Boston: Houghton Mifflin, 1955.

KATZ, D., MACCOBY, N., GURIN, G., & FLOOR, LUCRETIA. *Productivity, supervision, and morale among railroad workers.* Ann Arbor: University of Michigan, Survey Research Center, 1951.

MCGREGOR, D. *The human side of enterprise.* New York: McGraw-Hill, 1960.

MEUWESE, W. A. T., & OONK, SOPHIE. Enkele Determinanten van Creativiteit, Structuur, en Proces in Kleine Groepen. Unpublished master's thesis, University of Amsterdam, 1960.

MURRAY, H. A. *The Thematic Apperception Test.* Cambridge: Harvard Univer. Press, 1938.

PARNES, S. J., & MEADOW, A. Effects of "brainstorming" instructions on creative problem solving by trained and untrained subjects. *J. educ. Psychol.,* 1959, **50,** 171–176.

PEPINSKY, PAULINE N. Originality in group productivity: I. Productive interdependence in three natural situations. Columbus: Ohio State University, Research Foundation, 1959.

SPECTOR, D., & SUTTELL, BARBARA J. An experimental comparison of three patterns of leadership behavior. Washington, D. C.: American Institute for Research, 1956.

TAYLOR, D. W. Variables related to creativity and productivity among men in two research laboratories. In C. W. Taylor (Ed.), *The 1957 University of Utah Conference on the identification of creative talent.* Salt Lake City: Univer. Utah Press, 1958.

WEITZENHOFFER, A. M., & HILGARD, E. R. *Stanford Hypnotic Susceptibility Scale.* Palo Alto: Consulting Psychologists Press, 1959.

(Received August 3, 1961)

Journal of Abnormal and Social Psychology
1962, Vol. 65, No. 5, 319–324

DISSONANCE REDUCTION, SOCIAL COMPARISON, AND MODIFICATION OF OTHERS' OPINIONS

TIMOTHY C. BROCK AND JAMES E. BLACKWOOD[1]

Iowa State University *Pittsburgh, Pennsylvania*

Numerous investigations have shown that the stability, strength, and affective direction of opinions are primarily dependent on social reality, i.e., knowledge concerning the opinions of others. It follows that a person who changes his opinion may attempt to modify the social reality relevant to his opinion (others' opinions on the same topic). The present study concerned the generalization of a model of intrapersonal opinion change to account for changes in the perceived opinions of comparable others. An attempt was made to gather data on modification of others' opinions which could be ordered either to cognitive dissonance theory (Festinger, 1957) *or* to a theory of social comparison processes (Festinger, 1954) *but not to both*. Thus the relative potency of the two sets of principles in accounting for distortions of social reality could be assessed.

Studies conducted within the framework of dissonance theory (Festinger, 1957) have shown that an individual induced to take a stand at variance with his private opinion may subsequently show opinion change in the direction of that induced stand (see review by Cohen, 1960). Two recent studies (Cohen, Brehm, & Fleming, 1958; Rabbie, Brehm, & Cohen, 1959) showed that dissonance and consequent opinion change in line with the induced stand were greater, the less the justification provided for taking the discrepant stand. A straightforward extension to modification of others' opinions is possible: when a person is induced to take a stand contrary to his opinion, modification of comparable others' perceived opinions so that they agree with the stand taken will be greater, the less the justification offered. Of course, this effect would only be expected if others' opinions are the same as the person's original opinion. Where others' opinions are already in line with the induced stand, their modification should increase rather than reduce cognitive dissonance.

Derivation of the second hypothesis was guided by Festinger's (1954) theory of social comparison processes. Two of his propositions and an additional assumption were involved: subjective evaluations of opinions are stable when comparison is available with others who are judged to be close to one's opinions (p. 122); when a discrepancy exists with respect to opinions there will be tendencies to change others to bring them closer to oneself (p. 126); subjective evaluation of opinions is less stable and the effect of an opinion discrepancy is greater under high than under low magnitudes of dissonance. Thus when a person is induced to take a stand contrary to his opinion, the tendency to perceive others' opinions as closer to his own will be greater, the less the justification offered. The present experiment was designed to make two responses incompatible for the subject: changing others' opinions to bring them closer to his own original position (social comparison); changing others' opinions in the direction of his discrepant stand (dissonance reduction).

METHOD

An after-only design with appropriate control groups was employed. Subjects were told they were participating in a "survey" and were first asked to examine the "opinion data" collected from "previous respondents." Half the subjects were given "data" showing a large majority in favor, and half, a large majority opposed on an issue of highest importance to them. Then they were asked to write essays against their private opinion on the same issue. To create low dissonance in half of the subjects, the experimenter strongly justified writing the discrepant essay and to create high dissonance in the other half, the experimenter barely justified it.

[1] The comments of A. R. Cohen on a preliminary version of this report are gratefully acknowledged. The study was conducted while the authors were at the University of Pittsburgh.

The main dependent variable was subjects' recall of the opinions of "previous respondents."

Subjects

Subjects were 124 underclass students at the University of Pittsburgh who were randomly assigned to four experimental and three control conditions.

Procedure

The opinion issue, "increasing tuition by $10.00 per credit," was very important to all students and it was expected they would be strongly opposed. The student newspaper published a series of articles attacking a $4.00 per credit increase that was instituted several months before the experiment.

The experimenter gave the subjects reporting to him individually the following printed instructions:

> A survey of students' opinions on various subjects is being conducted at the University of Pittsburgh. One of the topics is *tuition*. In the past few months, a sizable number of students has been interviewed, and one of the questions concerned a raise in tuition of $10 per credit. Look through the accompanying pile of index cards to see if you think you know any of the students who have participated in the survey. Each card contains the following information: (*a*) first name and initials; (*b*) major; (*c*) opinion about raising tuition; (*d*) file number. Go through the cards carefully writing down the file numbers of the cards containing data from students that you are confident you know.

The subjects looked through 96 cards (3″ × 5″) and wrote the file numbers of cards containing data from students they thought they knew. The names and opinions on the cards were fictitious. In the Favorable Majority condition all but five of the cards contained either the words "favorable" or "very favorable," under the heading Opinion about Raising Tuition. In the Opposed Majority condition all but five of the cards contained either "opposed" or "very opposed." Thus in the Favorable Majority condition approximately 95% of others' opinions favored tuition increase; in the Opposed Majority condition, only 5% favored the increase.

When the subject completed the examination of the cards (in about 4 minutes), the experimenter gave him further printed instructions:

> In addition to collecting opinions some students have also been asked to write essays on the tuition topic. Enough essays taking a position opposed to a tuition increase have already been collected. To complete the survey, essays in favor of tuition increase are still needed. Write a *forceful, compelling,* and *persuasive* essay with the following title: "Why I Want the Tuition at the University of Pittsburgh Increased by $10 per Credit." Copy the essay title on the next page and write your name below it.

When subjects in the High Justification condition were about to copy the essay title, the experimenter intervened and gave reasons for writing the discrepant essay. This manipulation was essentially the same as in previous studies (Cohen et al., 1958; Rabbie et al., 1959).

> There are many good reasons for asking you to write in favor of a tuition increase, regardless of your own personal opinion. The ability to stand back and consider all sides of an issue is a real asset and advantage for any college student, for any mature person. By writing this essay you will help research in its scientific study of this new method as a way of arriving at solutions to age-old controversies such as those between labor and management or between nations in international affairs. This student opinion project is dependent on the cooperation of many students. Therefore, we are dependent on your cooperation. Finally, let me say that writing this essay will help me, help science, and help you in the sense that you take this opportunity to organize your own ideas and beliefs.

In the Low Justification condition the experimenter did not intervene and give reasons.

As soon as the subjects wrote the essay title and signed their names below it, the experimenter said:

> Before you actually go ahead on the essay, I'd like you to fill out this questionnaire. Please answer as accurately and truthfully as you can.

The first item measured subjects' private opinion toward the tuition increase of $10.00 per credit. Answers were indicated along a 71-point scale, each tenth point identified by labels running from Completely Opposed (0) to Completely Favorable (7.0). The second item, "Recall as best you can what percentage of the students listed on the cards you looked through were *in favor* of the tuition increase," was answered by checking a 101-point scale, each tenth point identified by labels running from 0% to 100%. A subject's score on this second item was the main dependent variable. (It must be noted that administration of the opinion and recall items preceded any actual work on the essay.) Following the questionnaire, subjects wrote for about 1 minute and then the experimenter said:

> We can stop now because we are mainly interested in the first thoughts on this topic that come to students' minds. But before you leave please tell me who pays your tuition bills.

Then the experimenter explained that how a student's tuition is financed may affect his opinion about an increase and that, as far as he knew, no increase was planned by the administration.

In addition to the four experimental conditions (Favorable and Opposed Majority, High and Low Justification) there were three control groups: subjects who were given the opinion measure only

(no experimental treatments);[2] subjects who were given only either the Favorable Majority or Opposed Majority treatment and the same questionnaire as the experimental subjects.

RESULTS

Effectiveness of the Experimental Manipulations

Acceptance of the fictitious majorities was similar in the four experimental and two relevant control conditions. There was no difference between the six groups in the average number (and range) of "previous respondents" whom the subjects indicated they were confident they knew (overall $M = 3.3$; range from 2 to 9). The "respondents" subjects said they knew were always members of the majority, e.g., in the Favorable Majority condition, the file numbers written by subjects were for others whose opinions were favorable to the tuition increase.

The favorable and opposed majorities had, in themselves, no effects on opinion scores. The means for these control groups (shown in the sixth and seventh rows of Table 1) were not different from the mean of the control group which only received the opinion measure (fifth row of Table 1). Opinion scores were not statistically greater in the Favorable than in the Opposed Majority conditions. However, since a difference might have been expected, this is discussed below.

The manipulation of justification produced differences consistent with theory and previous results. Opinion scores were higher in the Low than in the High Justification conditions (first four rows of Table 1). Tests of the differences between the Low Justification conditions, on the one hand, and the High Justification and control groups on the other, yielded t's greater than 3.00 ($p < .01$).[3] On the assumption that a more favorable opinion toward tuition increase would be produced under high than under low dissonance, it was concluded that the manipulation of justification effectively created high and low magnitudes of dissonance.

TABLE 1

OPINION SCORES IN EXPERIMENTAL AND CONTROL CONDITIONS

Condition	*N*	*M*	*SD*
Low Justification, Favorable Majority	17	2.24[a]	1.40
Low Justification, Opposed Majority	15	2.13	1.48
High Justification, Favorable Majority	15	.63	.36
High Justification, Opposed Majority	15	.77	.75
Control I (Opinion Measure only)	42	.71	.82
Control II (Favorable Majority only)	10	.66	.87
Control III (Opposed Majority only)	10	.59	.45

[a] The higher the score, the more positive is the attitude.

Recall of Others' Opinions

Completely veridical recall of the percentage of "previous respondents" favoring a tuition increase implied a report of about 95% in the Favorable and 5% in the Opposed Majority conditions. If dissonance theory predicts deviations from these baselines then, where dissonance was greater, there should be a greater tendency to modify others' opinions so that they agree with the discrepant stand (commitment to write the essay). Specifically, in the Opposed Majority condition, subjects should report considerably more than 5% of others' opinions as favoring the tuition increase. In the Favorable Majority condition, no substantial distortion of others' opinions would be expected since here the majority was represented as supporting the discrepant stand. In the latter condition, subjects might derive comfort from the knowledge that the contrary position they were about to support was supported by the overwhelming majority of previous respondents.

Principles of social comparison imply different outcomes. If dissonance increases a person's drive to evaluate the correctness of his opinion, and, if such an evaluation is facilitated by the proximity of others' opinions to his own, then, where dissonance is greater, there should be a greater tendency to modify others' opinions in the direction of his own. Specifically, in the Opposed Majority condition, subjects should report no substantial distortion of others' opinions since the opinions of subjects in the present

[2] The number of subjects in this control group was substantially larger ($N = 42$) than in the other conditions, because the data were collected as part of a questionnaire administration in two classrooms.

[3] All t tests in the present report were two-tailed. Where variances were heterogeneous an approximation technique by Edwards (1950, p. 168) was employed.

TABLE 2

MEAN PERCENTAGE OF OTHERS JUDGED AS FAVORING TUITION INCREASE

Condition	*N*	*M*	*SD*
Low Justification, Favorable Majority	17	52.88	28.46
Low Justification, Opposed Majority	15	17.47	11.87
High Justification, Favorable Majority	15	86.80	7.00
High Justification, Opposed Majority	15	12.80	9.60
Control II (Favorable Majority only)	10	83.90	8.94
Control III (Opposed Majority only)	10	12.70	7.44

study (in spite of differences produced by the justification manipulation) remained unfavorable to the tuition increase. In the Favorable Majority condition, subjects should report considerably less than 95% of others' opinions as favoring the tuition increase.

The results were consistent with social comparison theory (see Table 2). Under low dissonance (High Justification) deviations from the 95 and 5% baselines were the same as for the control groups. With the majority opposed, the mean percentage of others judged as favoring the tuition increase under high dissonance (Low Justification) was not statistically larger than the mean percentage under low dissonance (t of difference between 17.47 and $12.80 = 1.12$, $p > .25$). With the majority favorable, the mean percentage of others judged as favoring the tuition increase was much smaller under high than under low dissonance [4] (t of difference between 52.88 and $86.80 = 4.16$, $p < .001$).

An analysis of the product-moment correlations between attitude scores and recall deviations (from 95 or 5%, where appropriate) suggested that intrapersonal opinion change and modification of others' opinion were unrelated processes. For the four experimental conditions, as ordered in Table 2, the correlation coefficients ($-.02$, $+.08$, $+.28$, and $-.21$) were not different from zero and comparisons among them were nonsignificant.

DISCUSSION

It was noted that presentation of favorable and opposed majorities did not, per se, lead to differences in opinion scores (Table 1). This outcome may be compared to a previous failure to set up differential opinion change pressure merely by manipulating the size of fictitious majorities (Robbins, 1961). Majority opinion may be only effective when other factors are present, for example, face-to-face interaction (Deutsch & Gerard, 1955), attraction to a group (Festinger, Schachter, & Back, 1950), or unstructured stimulus situations (Sherif, 1936). The present subjects had no basis for attraction to the 96 persons whose "data" they examined; in fact, only a small percentage of "previous respondents" were considered to be persons they knew. Hence, the social reality of the present study consisted of persons who were comparable to the subjects only in that they were also students at the same university. In view of this minimal comparableness, it might be claimed that the hypotheses were tested in an especially stringent fashion. At least, it may be concluded that the direction of others' opinions on an issue was not critical until the person became embroiled in a conflict involving his own opinion on that issue.

A dissonance explanation of the present results may be considered. Since the mean opinion scores (2.2 and 2.1) for high dissonance subjects (Low Justification) remained unfavorable (the scale midpoint was 3.5) it might be alleged that undercalling the number in favor in the Favorable Majority condition served dissonance reduction by reinforcing the original opinion position. Subjects who did adopt a favorable opinion (at or across the midpoint) would increase dissonance by the undercalling. This account implied less undercalling, the closer subjects came to adopting a favorable opinion, or an inverse relation between magnitude of dissonance (as measured by opinion scores) and deviations from the 95% baseline. However, it was noted that the correlation between recall deviations and opinion scores in the Favorable Majority, Low Justification condition was almost zero ($-.02$). Thus minimal justification for taking a contrary position did not lead to distortions of others' opinions that could be readily interpreted in terms of dissonance reduction. Instead, a different process was apparently instigated: the person about to take a stand at variance with his own opinion modified others' opinions so that they concurred more with his

[4] See Footnote 3.

own prior opinion. It may tentatively be concluded that Festinger's (1954) earlier formulation provides a superior explanation of distortions of social reality. A person's need to perceive others as agreeing with his original opinion may be more powerful than a need to perceive others' opinions as supporting his discrepant behavior. However, the present study did not directly test an important intervening assumption: the strength of the drive to evaluate the correctness of a belief is directly related to the magnitude of dissonance aroused by commitment to behavior contrary to that belief. Confirmation of this hypothesis may be a useful step towards a desirable theoretical objective: resolution of dissonance and social comparison theories into a unitary set of assumptions and propositions.

The present results might be accounted for in a somewhat different fashion. High dissonance subjects (Low Justification), in contrast to the lows, were sensitized by the discrepant stand they were about to take so that they mobilized higher standards for dealing with opinion reality on the tuition increase. This led them to reject the information given on the file cards from previous respondents and to report more veridically what was known from their own experience, namely, widespread student opposition to tuition increases. When the prospective discrepant stand mobilized higher standards for assessing opinion reality, this accuracy motive might have overridden the possible increase in dissonance implied by undercalling the number in favor in the Favorable Majority condition.

The force of this explanation stems mainly from the plausible assumption that the subjects had considerable prior information about the opinions of others concerning a tuition increase. Replication of the experiment with a less reality-bound issue would be one means of eliminating the problem. It was possible, however, to evaluate this explanation in terms of the present data. First, the fact that the high dissonance subjects tended to overcall the number in favor in the Opposed Majority condition is not readily accounted for. Although this trend was not statistically significant, it was at least in a direction contrary to what might be expected from subjects who reminded themselves of student newspaper attacks on tuition increases and of the opposition of their friends. Second, it may be assumed that rejection of the file card information was more difficult for a subject who reported a greater number of previous respondents as being persons he "was confident" he "knew." If the explanation under consideration is valid, there should be an inverse relation between number of previous respondents listed as being known and deviations from the baselines in the Favorable and Opposed Majority conditions under high dissonance. Biserial correlations between these variables were close to zero. Third, the realistic orientation said to be instigated in the high dissonance subjects might be reflected in the arguments written in favor of the tuition increase (in the brief time allowed before the experiment was terminated). However, an analysis of the content of the sentence or two written by each subject did not show differences between the High and Low Justification conditions in the appropriateness of the reasons for a tuition increase, number of specific facts cited (e.g., new library being planned, eleventh floor not yet completed), reference to personal experiences (e.g., poor teaching from underpaid professor). Hence, the alternative explanation was not considered particularly compelling.

Opinions of the present subjects were measured before work on the discrepant essays began. Thus an incidental outcome was the demonstration that commitment to discrepant behavior is sufficient to produce cognitive dissonance. This result corroborates the same finding by Rabbie et al. (1959). A problem for subsequent research concerns specification of the conditions under which commitment alone and commitment plus behavior are necessary to produce dissonance.

SUMMARY

The present study concerned the generalization of a model of intrapersonal opinion change to account for changes in the perceived opinions of comparable others. Data was gathered which could be ordered either to cognitive dissonance theory (Festinger,

1957) or to a theory of social comparison processes (Festinger, 1954) but not to both.

Sixty-two underclass college students were told they were participating in a "survey" and were asked to examine the "opinion data" collected from "previous respondents." Half the subjects were given "data" showing a large majority in favor, and half, a large majority opposed to a tuition increase of $10.00 per credit. Then they were asked to write essays in favor of a tuition increase (they were initially very opposed). High and low dissonance was created by giving few and many justifications for taking the discrepant stand. The main dependent variable, measured immediately before the essays were actually written, was the subjects' recall of the opinions of previous respondents. Two responses were incompatible for the subject: changing others' opinions to bring them closer to his own prior position (social comparison); changing others' opinions in the direction of his discrepant stand (dissonance reduction).

Minimal justification for taking a contrary position did not lead to distortions in recall of others' opinions that could be interpreted in terms of dissonance reduction. Instead, the person about to take a stand at variance with his own position modified others' opinions so that they concurred more with his own. It was concluded that Festinger's (1954) earlier formulation provides a superior explanation of distortion of social reality. However, dissonance theory did account for differences in own opinion between the high and low justification conditions.

REFERENCES

COHEN, A. R. Attitudinal consequences of induced discrepancies between cognitions and behavior. *Publ. opin. Quart.*, 1960, **24**, 297–328.

COHEN, A. R., BREHM, J. W., & FLEMING, W. H. Attitude change and justification for compliance. *J. abnorm. soc. Psychol.*, 1958, **56**, 276–278.

DEUTSCH, M., & GERARD, H. B. A study of normative and informational social influences upon individual judgment. *J. abnorm. soc. Psychol.*, 1955, **51**, 629–636.

EDWARDS, A. L. *Experimental design in psychological research*. New York: Rinehart, 1950.

FESTINGER, L. A theory of social comparison processes. *Hum. Relat.*, 1954, **7**, 117–140.

FESTINGER, L. *Theory of cognitive dissonance*. Evanston, Ill.: Row, Peterson, 1957.

FESTINGER, L., SCHACHTER, S., & BACK, K. *Social pressures in informal groups*. New York: Harper, 1950.

RABBIE, J. M., BREHM, J. W., & COHEN, A. R. Verbalization and reactions to cognitive dissonance. *J. Pers.*, 1959, **27**, 407–417.

ROBBINS, P. R. Immediate and delayed effects of social influence upon individual opinion. *J. soc. Psychol.*, 1961, **53**, 159–167.

SHERIF, M. *The psychology of social norms*. New York: Harper, 1936.

(Received August 4, 1961)

Journal of Abnormal and Social Psychology
1962, Vol. 65, No. 5, 325–332

TEAM COMPETITION, SUCCESS, AND THE ADJUSTMENT OF GROUP MEMBERS [1]

ALBERT MYERS [2]
University of Illinois

There has been considerable speculation among educators, men of letters, and parents about the effect of competition upon interpersonal relationships and adjustment. The rational arguments for and against competition in schools, sports, or the military services are legion. In the foreground of consideration is the problem of the individual's adjustment to his group when more than one group is engaged in interteam competition. Specifically, what effect does intergroup competition have on the individual's psychological well-being and his interpersonal relations with others in the group of which he is a member?

The adjustment of individuals to their groups has been of primary interest in several recent studies by Fiedler and his associates (Alexander & Drucker, 1960; Fiedler, Hutchins, & Dodge, 1959; McGrath, 1962). They have been working on the problem of isolating the antecedent variables which facilitate good group member adjustment. The major emphasis has been upon the interpersonal perceptions of the group members. The present research is an attempt to relate certain situational variables, namely, competition and success, to the group adjustment problem.

Previous research on the effects of competition on adjustment has tended to focus on the competition-cooperation distinction. Thus, Deutsch (1949); Stendler, Damrin, and Haines (1951); Hirota (1951); Mizuhara and Tamai (1952); Blau (1954); Grossack (1954); and Gottheil (1955) all indicated that there is better understanding and positive affect between individuals cooperating with one another than between individuals competing against each other. That is to say, there are better relationships between teammates than between opponents.

In addition to the differential perceptions an individual has of teammates and opponents, it seems likely that an individual might have different perceptions of teammates in competitive and noncompetitive situations. A competitive situation is probably a more intense social experience than a noncompetitive situation. One important issue to be investigated is whether the intensity of a competitive experience between groups has any systematic effect on the adjustment and interpersonal relations of group members.

A competitive experience immediately implies the existence of potential success and failure. It seems likely that success should make it easier for individuals to interact with each other in a more harmonious fashion while failure should put a strain on interpersonal relationships. The first hypothesis, consequently, is that individuals in successful teams will tend to have more favorable interpersonal perceptions of their teammates than individuals on unsuccessful teams.

It seems possible, however, that success has a much different effect in a competitive situation than in a noncompetitive one. Since the goals are assumed to be more important in competition, success should be particularly satisfying and failure should be particularly frustrating. If the favorableness of the perceptions of teammates covaries with the degree to which the individual is rewarded, competition should lead to greater

[1] This investigation was supported by Contract Number DA-49-193-MD-2060, "Interpersonal Perception and the Psychological Adjustment of Group Members," between the Office of The Surgeon General, Department of the Army, and the University of Illinois (Fred E. Fiedler, principal investigator).

Fred E. Fiedler, the author's thesis adviser, and Joseph E. McGrath aided in the design and data collection phase of this experiment. James W. Julian and Scott E. Krueger assisted during data collection and analysis. The writer is especially appreciative of the cooperation which he received from Clair Worthy, Commander of the University of Illinois Army ROTC unit, and his staff. In particular, D. McClelland, W. Hooper, and M. Snoke helped to coordinate the project. E. F. Flanders served as consultant and range safety officer.

[2] Now at the Educational Testing Service.

variability of interpersonal perception than noncompetition.

The second hypothesis, therefore, is that success in a competitive situation will generate more favorable perceptions of teammates than success in noncompetitive situations. The third hypothesis is that failure in a competitive situation will tend to generate more unfavorable perceptions of teammates than failure in noncompetitive situations.

METHOD AND PROCEDURE

Subjects

A Men's Recreational Rifle Tournament was organized with the assistance of the Army Reserve Officer Training Corps (ROTC) unit at the University of Illinois. Announcements of the tournament were made to freshman and sophomore ROTC students during their regular class period. Of approximately 430 students who registered during the following week, 180 were selected.

Design

Sixty teams, each consisting of three men, were formed. Half of these teams were combined into competitive leagues and half into noncompetitive leagues. There were several major differences in the manner in which the two types of leagues were conducted. All of these were designed to emphasize salient aspects of the competitive and noncompetitive conditions while keeping the information obtained by the subjects, their prospect of success, and the objective reward constant. No individual scores were made available to the subjects. Only team results were reported.

Competitive Leagues. Five separate competitive leagues composed of six teams each were formed. The leagues met 1 hour a week for 5 weeks. The first 2 weeks were devoted to marksmanship, safety instructions, and practice firing. Starting with the third week, each league began a 25-bout round robin schedule. Win-loss records were kept and announced for each team and members of the winning team in each league were promised individual trophies.

In order to facilitate the interteam evaluative aspect of this condition, the experimenter announced the results of each bout and made frequent references to the fact that certain teams were doing well and others badly.

Noncompetitive Leagues. The main experimental purpose in the noncompetitive leagues was to prevent interteam comparisons (i.e., competition). The teams did not compete against each other but against a set of three qualification standards of hard, moderate, and easy difficulty. The teams knew in advance against which standard they were firing on each bout (this is the analogous information of knowing how good your opponent is in the competitive condition). The experimenter announced the results of each bout quietly to each team and never made evaluative statements as to a team's past performance. The subjects were told that all members of teams which "qualified" would receive trophies. Although "qualification" was undefined specifically, the subjects were told that the performance standards were appropriate to their experience. It was emphasized that any number, or none of the teams might win trophies.

The first 2 weeks of practice firing and orientation were identical to the procedure used for the competitive condition.

Absences

In order to create some pressure for good attendance, penalty scores amounting to 20% of an absent member's previous average were added to the team score.

Measures

Three different types of measures were collected to assess various aspects of the individual's relationship to his team. They were: The individual's esteem for his teammates, the individual's perceived acceptance of himself by the team, and the individual's perception of who or what was to blame on those bouts on which the team failed.

Esteem for teammates was obtained on scales of the semantic differential type with items used by Fiedler in his work. The subjects were asked to rate each of their teammates on 24, six-point scales. The items were of the form friendly-unfriendly, cooperative-uncooperative. Esteem for a particular teammate was determined by simply summing the assigned scores on the 24 items.

Perceived Acceptance scores were determined as the sum of the subjects' responses to five, 16-point questionnaire items which were designed to assess how comfortable and accepted the subject felt in his group.

A multiple-choice item was used on which each subject indicated what he believed to be the reason for failure on those bouts in which his team was unsuccessful. The items reflected extrapunitive, intrapunitive, or impunitive modes of response. For example, he could be extrapunitive by saying that one of his teammates did poorly, or he could be intrapunitive by saying that he did poorly himself, or he could be impunitive by saying that the other team performed very well.

RESULTS

Esteem for Teammates

A subject's esteem to his group was measured by the average esteem he had for his teammates. A team score for esteem was obtained by averaging the individual scores. Thus, a high team esteem score required that all three members describe each other favor-

ably. Team esteem scores in the intermediate range could be achieved by innumerable combinations of high, medium, and low scores.

The means for team esteem scores in the second and fifth weeks and for the changes between these weeks are presented in Table 1. An analysis of variance on the change scores summarized in Table 2 showed that there was a significant condition effect ($p < .05$), i.e., the competitive teams showed more improvement in team esteem than did the noncompetitive teams. In addition, the high success teams showed an improvement in average esteem during the course of the tournament while the low success teams deteriorated. The difference of these changes was significant at the .01 level. The effect of success was also significant when each condition was considered separately. Thus, both the independent variables had a significant effect on the team esteem scores.

These data do not support the hypothesis that the competitive condition should generate greater variability between high and low success teams. The difference between the high and low teams was 10.4 ($p < .025$) in the competitive conditions and 16.9 ($p < .001$) in the noncompetitive condition.

TABLE 1

MEAN ESTEEM FOR WEEKS 2 AND 5 AND MEAN CHANGE BETWEEN WEEKS

Condition	*N*	Week 2	*N*	Week 5	*N*	Change[a]
Competitive level of success						
H	10	107.5	11	118.7[b]	10	9.1
M	7	113.0	7	112.0	7	−1.0
L	11	105.8	11	104.5	11	−1.3
Average	28	108.2	29	111.7	28	2.5
Noncompetitive level of success						
H	10	107.1	10	110.5	10	3.4
M	7	106.0	7	103.3	7	−2.7
L	10	107.7	12	100.0[b]	10	−11.7[b]
Average	27	107.0	29	104.4	27	−4.4

[a] Difference in *N* and rounding errors account for minor discrepancies in summing and averaging.

[b] Data from three teams were unavailable in Week 2. The change scores are based only on teams for which there were data in Week 5 and Week 2. Two teams, which averaged 120 in the noncompetitive condition with low success were among those missing in Week 2. Their scores are included in the Week 5 data. If they are removed, the low success average drops from 100.0 to 96.0 and the noncompetitive condition average drops to 103.2.

TABLE 2

SUMMARY OF ANALYSIS OF VARIANCE OF ESTEEM CHANGE SCORES

Source	*df*	*MS*	*F*
Conditions (C)	1	60.0	4.95*
Success (S)	2	83.0	6.85**
C × S	2	6.0	
Error	49	12.1	
Total	54		

* Significant at .05 level.
** Significant at .01 level.

The interaction *t* test to determine whether there was significantly more variability in esteem scores in the noncompetitive condition was insignificant. Thus, success had a uniform effect on the two conditions. This finding does not suggest that large discrepancies exist between the conditions in the subjective value of success. This notion is supported by data obtained from a questionnaire item. The responses of subjects in the noncompetitive condition indicated that they enjoyed the activity no less than did subjects in the competitive condition.

It is also somewhat unlikely that there were great differences in the subjective probability of success between the two conditions. The teams doing poorly, in both conditions, realized that their chances for a trophy were poor while the better teams generally felt that they had a good chance. Even in the noncompetitive condition where the standard was ambiguous, the trophy oriented attitudes seemed appropriate. It seems, therefore, that the differences in esteem cannot be attributed to differences in task interest or subjective probability of success.

Analysis of the Week 5 data alone showed, as would be expected, the same relationships found in the change data. Both the condition and success variables were significant at the .05 level. In addition, the successful competitive teams had significantly higher scores ($p < .05$) than the successful noncompetitive teams. Contrary to Hypothesis 3, however, the unsuccessful competitive teams had higher esteem scores than the unsuccessful noncompetitive teams. This difference, however, was not significant.

TABLE 3

MEAN ESTEEM GIVEN IN WEEK 5 TO PRESENT MEMBERS AND ABSENT MEMBERS FOR TEAMS WITH ABSENTEES IN WEEK 5

Condition	Present in Week 5	Absent in Week 5	Average
Noncompetitive	112.6	92.8	102.7
Competitive	108.3	109.3	108.8

Esteem for Absent and Present Members

The two conditions did not differ in the number of absences; 15 subjects in the competitive condition and 16 subjects in the noncompetitive condition were absent in Week 5. This represented an absentee ratio of 17.2% and 18.3%, respectively. Similarly, 12 competitive teams and 11 noncompetitive teams had no absences throughout the entire tournament.

There were 12 teams in each condition with absent members in the last week. If a team had one man absent, each of the present members had an opportunity to rate one absent teammate and one present teammate. Table 3 shows the distribution of scores for present and absent members in each condition. Whereas there was virtually no difference between the esteem scores given to absent and present members in the competitive condition, substantial differences were obtained in the noncompetitive condition. In the noncompetitive groups, the present members received an average esteem score of 112.6 while the absent members received a score of 92.8. This difference of 19.8 points had a two-tailed significance of .02.

The finding raised the question whether these subjects failed to attend in Week 5 because they were not liked in Week 4 or whether they were liked less in Week 5 because they had been absent in Week 5. The data indicated that those individuals who were absent in Week 5 received an average score of 102.1 in Week 4 and those who were present in Week 5 received an average score of 105.5. The difference between these scores was clearly not significant ($t = .48$). The relative change in esteem from Week 4 to Week 5 was, however, significant at the .05 level by the Wald-Walfowitz run test. There was, incidentally, no overlap in the distribution. None of the absentees received higher esteem scores in Week 5 while none of the present members received lower scores. The data, thus, indicated that the subjects who were absent in Week 5 in the noncompetitive condition were given low scores because they were absent in Week 5.

Perceived Acceptance

The esteem data were based upon a monadic judgment by each subject, i.e., the degree to which he liked each of his teammates individually. The five-item questionnaire (Perceived Acceptance) attempted to assess each subject's perception of the dyadic relationship that existed for a subject and his team, considered collectively. The two measures were not, of course, unrelated. They were meant to be two aspects of an individual's adjustment to a group. The correlation between these measures was .60.

Because certain items on the questionnaire were irrelevant until the competition began, the questionnaire was not administered until the third week.

The means for the Week 3 and the Week 5 data and the mean differences between weeks are shown in Table 4. As with esteem, the competitive teams tended to score higher (more favorably) than the noncompetitive teams and high success tended to lead to more favorable scores than low success.

TABLE 4

MEAN PERCEIVED ACCEPTANCE IN WEEKS 3 AND 5 AND MEAN CHANGE BETWEEN WEEKS FOR ENTIRE SAMPLE

Condition	*N*	Week 3	*N*	Week 5	*N*	Change
Competitive level of of success						
H	11	58.4	11	64.5	11	6.1
M	7	59.9	7	61.0	7	1.1
L	11	52.6	11	55.2	11	2.4
Average	29	56.7	29	60.1	29	3.4
Noncompetitive level of success						
H	10	55.4	10	55.9	10	0.5
M	7	55.2	7	51.6	7	−3.6
L	12	51.9	12	55.1	12	3.2
Average	29	53.9	29	54.5	29	0.6

An analysis of variance of the Week 5 data, summarized in Table 5, indicated a very strong condition effect. The mean difference of 5.6 in Perceived Acceptance scores between the competitive and noncompetitive teams was significant at the .01 level, but there was no overall effect due to success. The failure to achieve an overall success effect was obviously attributable to the noncompetitive condition which had very little variation. The mean difference between high and low success teams in the noncompetitive condition was only .8.

Success had, however, a significant effect within the competitive condition. The 9.3 difference between the high and low success group means on Perceived Acceptance scores was significant beyond the .005 level ($t = 3.10$). The high success competitive groups also scored more favorably than the high success noncompetitive teams ($t = 3.18$, $p < .001$). It should be noted, however, that the larger variability in the competitive condition was not due to poor adjustment in unsuccessful teams. The unsuccessful teams had a higher average score than the entire noncompetitive condition. Rather, the variability must be attributed to the extremely favorable scores obtained from the successful competitive teams.

Thus, the second hypothesis, which stated that competitive teams of high success would have more favorable adjustment scores than the noncompetitive teams of high success, was substantiated. The data relating to the third hypothesis, on the other hand, were once again, in the opposite direction. Instead of having lower scores than the noncompetitive teams, the low success competitive teams tended to have higher scores.

TABLE 5

SUMMARY OF ANALYSIS OF VARIANCE FOR PERCEIVED ACCEPTANCE IN WEEK 5 FOR ENTIRE SAMPLE

Source	*df*	*MS*	*F*
Conditions (C)	1	55	9.65**
Success (S)	2	14	
C × S	2	13.5	
Error	52	5.7	
Total	57		

** Significant at .01 level.

TABLE 6

DISTRIBUTION OF CHOICES FOR PUNITIVE QUESTION IN WEEKS 3 AND 5

Week	Modes of response	Competitive	Noncompetitive
First week of competition (Week 3)	Extrapunitive A (teammate did poorly)	28%	21%
	Extrapunitive B (poor equipment)	03	03
	Impunitive A (other team performed well; standards too difficult)	22	28
	Impunitive B (bad luck)	12	10
	Intrapunitive (I did poorly)	35	38
		100%	100%
Last week of competition (Week 5)	Extrapunitive A	21%	35%
	Extrapunitive B	00	05
	Impunitive A	35	18
	Impunitive B	10	17
	Intrapunitive	34	25
		100%	100%

Thus, it seems that the competitive group members tended to perceive a far more favorable relationship between their teammates and themselves, while the noncompetitive groups apparently never generated a feeling of acceptance. This is in spite of the fact that the individuals had been together for several weeks and at least some of them had enjoyed success in the task. Although many grew to have high esteem for their teammates (particularly in the high success teams), there seemed to be a surprising inability for these individuals to achieve well-integrated interpersonal relationships. It is interesting to note that the final level achieved by the noncompetitive groups was roughly equal to the precompetition level found for both conditions. Perhaps the interpersonal relationships of noncompetitive group members can best be summarized by saying that they remained strangers to each other.

Punitiveness

Table 6 shows the distribution of punitive choices in Weeks 3 and 5. In Week 3 the distribution of the two conditions were virtually identical. The χ^2 testing the difference between these distributions was only 1.4 which is clearly not significant.

There was a significant difference between the conditions in Week 5 on the use of an item reading "one of my teammates did

poorly." This is a personal form of extrapunitiveness. As Table 6 shows under the entry Extrapunitive A, 21% of the competitive subjects chose this item while 35% of the noncompetitive subjects chose it. The critical ratio of 2.05 was significant at the .05 level (Walker & Lev, 1953).

This result indicates that the competitive teams were less extrapunitive. The second extrapunitive item (B, poor equipment) suggested the same conclusion, but had too low a response rate for an adequate test. Combining the two extrapunitive items yielded a difference which was significant at the .02 level ($CR = 2.38$). The noncompetitive groups were more extrapunitive in general and more extrapunitive towards their teammates in particular.

The competitive subjects chose Impunitive A ("the other team performed very well") more frequently than the noncompetitive subjects. There was, on the other hand, a tendency for the noncompetitive subjects to choose Impunitive B (bad luck) more frequently. Consequently, there was no difference between the conditions in the degree to which they were impunitive when the Impunitive A and Impunitive B scores were added together to form a single impunitive score.

As might be expected from the discussion of the individual items, the simultaneous analysis of all items resulted in a chi square of 8.06 which was significant at the .05 level. It was necessary to use only four categories since Extrapunitive B (poor equipment) received a total of three votes. It would be legitimate to add these scores to the Extrapunitive A (teammate did poorly) totals or to drop them completely. The latter method was used here in accord with current statistical theory.

Unsuccessful teams had, of course, many more experiences of failure than successful teams. It might be expected that the amount of extrapunitiveness would increase with the number of failures a team experienced. The point biserial correlation between these variables was, however, only .07. Thus, somewhat surprisingly, the amount of extrapunitiveness was essentially a condition effect rather than a function of a team's failure experiences.

DISCUSSION

Interpersonal Perceptions and Success

The data clearly demonstrated that an individual's interpersonal perceptions of his group members is significantly affected by the level of success enjoyed by the group. It is relatively simple for people to be friendly and happy when the situation is rewarding. When, on the other hand, the situation creates disappointment for the individuals, it is presumably far more difficult to maintain harmonious group relations. The present data did not, of course, indicate that the individuals on unsuccessful teams actually behaved in a less friendly manner. It is likely, however, that the less favorable interpersonal perceptions reflect, at least in part, antagonistic interpersonal behavior. It is also possible, although less likely, that the behavior of the subjects on unsuccessful teams was as friendly but that the perceptions of these behaviors were drastically distorted. Additional research is needed to answer this question.

Interpersonal Perception and Competition

It was not expected that competition would generate more favorable interpersonal perceptions than noncompetition. Hypothesis 3 predicted that the unsuccessful competitive teams would have very low average Esteem for Teammates and Perceived Acceptance scores. However, the unsuccessful competitive teams tended to have higher scores than the unsuccessful noncompetitive teams. Competition, consequently, led to better team relationships regardless of the level of success.

Adjustment

The major concern of the present study was with the adjustment of an individual to his group. The criteria were measures of interpersonal perceptions. Good group adjustment should imply more than a general satisfaction with teammates. It should also include an ability of team members to withstand disruptive influences; an ability to function adequately in the face of adversity. These properties are incorporated in Hemphill's (1956) concept of group viscidity, Gross and Martin's (1952) concept of co-

hesion, and Fiedler et al. (1959) concept of quasi-therapeutic relations.

As indicated in the presentation of the results, the effects of two forms of adversity could be estimated. They were absences and failures. Good adjustment should manifest itself by providing the individuals of a group with the capability of maintaining harmonious group feelings in the face of adversity. The well-adjusted group should not disintegrate or deteriorate in the face of disruptive influences.

The data showed that more deterioration occurred in the noncompetitive condition. Absent members were rated in a significantly less favorable manner and there was significantly more extrapunitive behavior among the noncompetitive teams. These data are interpreted as an indication of poor group adjustment. The subjects in the competitive condition, on the other hand, did not respond intolerantly towards absentees, nor were they very extrapunitive.

In addition, of course, the competitive teams had more favorable scores on the Esteem for Teammates and Perceived Acceptance measures. In fact, the competitive condition had more favorable scores than the noncompetitive condition on every measure and in every situation.

Thus, the conclusion that the members of competitive groups were better adjusted seems valid for the following reasons: They esteemed their teammates more highly, they perceived the relationships between their teammates and themselves as being more favorable, they were more tolerant of teammates who posed an objective threat to success through absenteeism, and they were less extrapunitive towards teammates. That is, they did not hold their teammates responsible for failure.

Competition seemed to have acted as a prophylactic against poor group adjustment. At the same time, it permitted extremely good adjustment for members of groups which enjoyed success. Competition had the important property of drawing teammates together. This unification is commonly described as a reaction to a common enemy. It is possible that a *common evaluator* will provide the same function. Individuals will unite when they are being evaluated as a group.

The existence of an outside evaluator perhaps forces the individual to evaluate all performance on which he feels he is being judged. His individual performance, however, is to some degree confounded with that of the group. The individual must evaluate himself as part of the group and any evaluation of the group is, to some extent, an evaluation of himself. If it is assumed that the individual tries to evaluate himself as favorably as possible then there should be a tendency to evaluate all those elements to which he is connected in a favorable manner also. This relationship is frequently seen in situations where people defend the actions of their parents, church, or country. It is not only that they identify themselves with these groups, but, in addition, that an outsider identified the individual as being a member of that group. The contingent relationship which exists between an individual and his group should create pressure on the individual towards perceiving his fellow group members favorably. In addition, this pressure should increase as the evaluation becomes more important to the individual.

It was previously mentioned that this study attempted to relate certain situational variables to Fiedler et al. (1959) work on adjustment in groups. Fiedler et al. have incorporated the concept of quasi-therapeutic relations to refer to those sets of interpersonal behaviors which have a facilitative effect on group adjustment. McGrath (1962) has pointed out that a major portion of these facilitative "behaviors" as judged by teammates are perceptually rather than behaviorally determined.

The present study indicated that the perceptions of individuals could be systematically altered so as to produce groups with greater interteam adjustment. In Fiedler et al. (1959) language, competition and success had the property of generating quasi-therapeutic relations between teammates. Then, such so-called quasi-therapeutic relations within groups appear to be a consequence of the group's relation to its external environment, as well as of the interpersonal dynamics within the group itself.

SUMMARY

Sixty three-man teams comprised of 180 ROTC students at the University of Illinois participated in a Recreational Rifle Tournament. Six teams in each of five leagues engaged in a face-to-face, 25-bout, round robin. Thirty other teams fired not against the performance of other teams but for qualification against standards. The first 2 weeks in the 5-week study were devoted to practice firing.

Adjustment of each man to his group was measured by a five-item scale and by scores of esteem for teammates on semantic differential scales. Multiple-choice questions yielded scores of extrapunitive, intrapunitive, and impunitive modes of response. These measures were taken at every firing session.

The findings indicated that better team adjustment was generated by the competitive experience and success in the task. It was also found, contrary to one hypothesis, that the low success competitive teams tended to manifest better adjustment than the low success noncompetitive teams. It was concluded that the competitive experience not only engendered good adjustment under favorable conditions (success) but that it acted as a prophylactic against poor adjustment in unfavorable situations (failure).

Two types of adversity could be identified in the experiment. They were absence of teammates, which caused teams to lose points in the competition, and team failure. In both these situations the members of competitive teams responded in a more tolerant and adaptive manner than did members of noncompetitive teams. This tolerance and the general tendency to have more favorable interpersonal perceptions seemed to justify the conclusions that the competitive condition generated better adjustment of individuals in their teams.

REFERENCES

ALEXANDER, S., & DRUCKER, E. H. The effects of experimentally modified interpersonal perceptions on social behavior and adjustment. Technical Report No. 9, 1960, University of Illinois, Group Effectiveness Research Laboratory, Contract DA-49-193-MD-2060, Project on interpersonal perception and the psychological adjustment of group members, Department of the Army. (Mimeo)

BLAU, P. Cooperation and competition in a bureaucracy. *Amer. J. Sociol.*, 1954, **59**, 530–535.

DEUTSCH, M. An experimental study of cooperation and competition upon group processes. *Hum. Relat.*, 1949, **2**, 199–232.

FIEDLER, F. E., HUTCHINS, E. B., & DODGE, JOAN S. Quasi-therapeutic relations in small college and military groups. *Psychol. Monogr.*, 1959, **73**(3, Whole No. 473).

GOTTHEIL, E. Changes in social perceptions contingent upon competing or cooperating. *Sociometry*, 1955, **18**, 132–137.

GROSS, N., & MARTIN, W. E. On group cohesiveness. *Amer. J. Sociol.*, 1952, **57**, 546–554.

GROSSACK, M. M. Some effects of cooperation and competition upon small group behavior. *J. abnorm. soc. Psychol.*, 1954, **49**, 341–348.

HEMPHILL, J. K. Group dimensions: A manual for their measurement. *Ohio St. U. Stud. Bur. Bus. Res. Monogr.*, 1956, No. 87.

HIROTA, K. Experimental studies of competition. *Jap. J. Psychol.*, 1951, **21**(3/4), 70–81.

MCGRATH, J. E. The influence of positive interpersonal relations on adjustment and effectiveness in rifle teams. *J. abnorm. soc. Psychol.*, 1962, **65**, 365–375.

MIZUHARA, T., & TAMAI, S. Experimental studies of cooperation and competition. *Jap. J. Psychol.*, 1952, **22**, 124–127.

STENDLER, C., DAMRIN, D., & HAINES, A. C. Studies in cooperation and competition: I. The effects of working for group and individual rewards on the social climate of children's groups. *J. genet. Psychol.*, 1951, **79**, 173–197.

WALKER, HELEN M., & LEV, J. *Statistical inference.* New York: Holt, 1953.

(Received August 7, 1961)

Journal of Abnormal and Social Psychology
1962, Vol. 65, No. 5, 333–337

PERSONALITY STYLE AND PERCEPTUAL REACTIVITY TO THE IMMEDIATE ENVIRONMENT[1]

ARNOLD MILLER AND ELEANOR KEMP[2]

Montana State University

This study investigates the relationship between certain consistent personality styles and perceptual awareness of the immediate environment. The personality styles studied derive from findings by Allport and Vernon (1933), which indicate a general tendency among people to behave in an expansive or restrictive manner. They find, for example, that individuals who gesticulate broadly also tend to be those who write large, etc., while those who are restrictive gesturally tend to be contractile in other areas of expression. Eisenberg (1937), repeating many of the measures used by Allport and Vernon, also confirms such behavioral styles.

The present study views style as a pervasive and nonvolitional tendency of the organism to interact with the immediate environment in one way rather than another. Applied to visual perception, pervasiveness of style suggests that an "expansive" person should view his immediate environment in a broad, ranging manner paralleling broad gestures, expansive speech, etc., while the "restrictive" person should restrict his view of the immediate environment just as he limits his gesture and speech. The nonvolitional aspect of style implies that it is a stable characteristic which does not vary with changes in a person's purposefulness. Thus, if a person's style entails broad scanning of his environment, this tendency should be evident whether he is purposefully or casually viewing particular surroundings. The existence of such pervasive and nonvolitional styles may be tested by observing how people perceive their immediate surroundings. But first it is necessary to consider the nature of these surroundings.

[1] This study was supported in part by Research Grant M 4915 from the National Institute of Mental Health, United States Public Health Service.
[2] The authors wish to thank Baldwin Hergenhahn for his help in conducting this study.

In an environment there are usually present both more and less highly valenced objects. Objects are thought to be more highly valenced and thus more easily noted if they serve a personal function, are near at hand, or are meaningfully grouped. In contrast, less highly valanced objects are thought to be easily ignored because they lack a personal function, are distant from the observer, or are background for more attractive items. The highly valenced items, in this study, are referred to as figure or figural objects, the less highly valenced objects as ground or ground-objects.

The following hypotheses anticipate the manner in which people will reflect their personality styles in the perception of such aspects of their immediate environment:

1. Expansive people are expected to show their style perceptually by noting significantly more aspects of their immediate environment than restrictive people.

2. Expansive people are expected to note relatively more ground and fewer figural objects than restrictive people; conversely, the latter are expected to note relatively more figural and fewer ground-objects than the former.

3. Both kinds of people are expected to manifest their styles independent of volitional changes in their attitude toward what they perceive.

METHOD

There were two phases to this study: the first was designed to permit selection of expansive and restrictive groups of people on the basis of their observed behavior; the second, a week later, tested these groups on their perceptual reactivity to a prepared environment.

Phase I

Subjects and Procedure. The subjects who went through this phase were 62 men and women students (30 men and 32 women) from courses in introductory psychology. Each subject was informed

that the study was concerned with how well people, without advance preparation, could express themselves. The subject was then led into a room arranged for one-way observation and tape recording. A man and woman, in their midtwenties, instructed the subject to talk to them on any subject or subjects of interest to him for 5 minutes. The subject was then placed in position 6 feet from and facing the seated "audience" with his body profile toward the vision screen. During the subject's discourse the audience was attentive but did not initiate comments unless the subject remained silent for more than 30 seconds. When initiated, comments were supportive and directed toward getting the subject talking again.

Measures. Five measures were obtained for each subject. These were total gestures, outward gesture proportion, total words, modifier proportion, and rating on a four-point expansive-restrictive (E-R) scale.

Total gestures. Use of many gestures was thought to be related to the expansive and few gestures to the restrictive style. Any accented movement was considered a gesture. Thus, each gesture made by any portion of the body during the subject's 5-minute presentation was scored by the observer behind the vision screen as well as by both members of the audience. Scored gestures included movements of the head (nods), of the shoulders (shrugs), of the entire torso (bowing, leaning forward), as well as movements of hands and arms. Any change in hand orientation was scored as another gesture. The subject's total gesture score was based on the mean of the three independent counts obtained during the 5-minute interval. Interobserver agreement for the three scores using Kendall's W was .70.

Outward gesture proportion. High proportions of outward gesture were assumed related to expansive and low proportions to restrictive personality styles. Vector of gestures was scored by the observer behind the one-way screen. The observer had three telegraph keys: one for recording the number of outward gestures, the other for inward gestures, and the third for neutral gestures that had no apparent vector. The outward gesture score was obtained by computing the proportion of outward gestures to the total gesture count.

Total words. It was assumed that high word counts were related to the expansive style and low counts to the restrictive style. Word counts for each subject were obtained by means of a concealed microphone linked to a tape recorder in the observation room. The mean of three independent counts of the subject's 5-minute word production became the subject's total word score. Interobserver agreement using Kendall's W was .82.[3]

Modifier proportion. High proportions of modifiers were considered related to expansive and low proportions to the restrictive style. The proportion of modifiers used in each subject's 5-minute word production, i.e., the number of adjectives and adverbs used, as determined by the mean of two independent counts, became the subject's score. The rho for the two independent counts was .60.

E-R rating scale. The purpose of this four-point scale was to estimate qualitative aspects of the subjects' behavior which did not lend themselves to tallying. If, during his talk, a subject covered a broad range of topics, spoke with marked intonation, and related easily, e.g., smiling, eye contact with audience, etc., he was scored highly expansive (4) on the rating scale. If he spoke on one topic in a monotonous voice and related with difficulty, e.g., frowning, blushing, little eye contact, etc., he was scored highly restrictive (1) on the rating scale; scores of 2 and 3 referred, respectively, to moderate restrictiveness and expansiveness. All subjects were independently rated by each of the three observers. Interobserver agreement of ratings computed with Kendall's coefficient of concordance was $W = .60$.

After the five scores were obtained, expansive and restrictive groups of people were established. This was done by converting all five scores for each subject into ranks based on the position of each score within its respective distribution, creating one mean rank score for each subject based on his five ranked scores, and ranking these 62 mean ranks. These subjects falling within mean ranks above the median (1–31) were designated expansive; those below (32–62) restrictive.

In order to examine the presumed consistency of the behavior style on which the classification procedure was based, Kendall's coefficient of concordance was computed for the five ranked scores. The $W = .34$ was significant at $p < .05$. Since the Kendall coefficient could not indicate how the five scores contributed to the W, Spearman's rank-order correlation (rho) was computed for all possible pairs of the five scores.

TABLE 1

RANK-ORDER CORRELATION COEFFICIENTS (rho) BETWEEN ALL POSSIBLE PAIRS OF FIVE SELECTION MEASURES

Measure	Rho
Total gestures vs. Outward gestures	.34***
Total gestures vs. Total words	.26**
Total gestures vs. Modifier proportion	−.03
Total gestures vs. E-R scale	.59***
Outward gestures vs. Total words	.20*
Outward gestures vs. Modifier proportion	−.07
Outward gestures vs. E-R scale	.38***
Total words vs. Modifier proportion	−.09
Total words vs. E-R scale	.41***
Modifier proportion vs. E-R scale	−.08

Note.—All significance levels are based on one-tailed tests.
* $p < .10$.
** $p < .05$.
*** $p < .01$.

[3] The lack of a perfect correlation between observers on this measure was due to difficulty in hearing distinctly the voices of some subjects.

Table 1 supported the existence of consistent personality styles across different behaviors. Only one of the five measures, modifier proportion, was inconsistent with the general tendency for the measures to intercorrelate substantially. The small negative correlations found between the modifier proportion and the other four measures, although not significant, suggested that a larger coefficient of concordance would have been found if this measure had not been used. Results also showed that the E-R scale contributed most toward the coefficient of concordance, correlating significantly with all other measures. An estimate of how much the E-R scale contributed to the classification of people into the two styles was obtained by computing the correlation between the ranked E-R scale scores and the mean ranks derived from all five ranked scores. A rho of .77 was found and suggested that a classification of subjects, based solely on the E-R scale, would be quite similar[4] to the classification based on the mean of all five ranked scores.

Phase II

Of the original 62 subjects completing Phase I, 44 returned a week later to take part in Phase II. These 44 were composed of 20 previously scored expansive (13 women, 7 men) and 24 previously scored restrictive (11 women, 13 men).

Testing for environmental reactivity. Under the pretext that the experiment was not ready to begin, each subject was required to "wait" 4 minutes in a room designed to give the general impression that someone in the process of setting up their office had been called away. This provided a rationale for having a variety of objects in the room, some arranged in fairly coherent groups, others scattered around the room.

The objects in the prepared room were categorized into figure and ground objects. For an object to be classified figure it had to be close to where the subject was sitting, have an easily discerned function suggesting manipulation, and be part of a group of related objects. For an object to be classified ground it had to be further from the subject than the figure objects, a structural part of the room, or placed haphazardly on the floor or window sill. Examples of one group of figure objects (arranged on the desk) included a typewriter, ruler, pencil, paper clips, matches, ash tray, etc.; ground objects included parts of the room such as the green wall, brown floor, or unrelated objects (two sticks of wood, empty cigarette carton, ventilator grill, etc.) placed at different points around the room. There were 62 items in the room, 31 previously judged figure and 31 previously judged ground.

[4] When classification of people into the two style categories was based solely on the E-R ratings, 51 out of 62 (82%) of the placements agreed with the classification based on the mean ranks. Thus, only 11 of the 62 placements (18%) were based on the other four measures.

TABLE 2

ANALYSIS OF VARIANCE FOR PERSONALITY STYLES AND PERCEPTUAL REACTIVITY

Source[a]	*df*	*MS*	*F*
c Figure-Ground (F-G)	1	406.66	67.88****
b Casual-Directed (C-D)	1	3638.81	257.70****
a Expansive-Restrictive (E-R)	1	23.51	<1.00
b F-G × C-D	1	91.26	6.46**
c E-R × F-G	1	29.51	4.92**
b E-R × C-D	1	4.31	<1.00
d E-R × C-D × F-G	1	.29	<1.00
a Subjects (*Ss*)	38	27.53	
b *Ss* × C-D	38	14.12	
c *Ss* × F-G	38	5.99	
d *Ss* × F-G × C-D	38	6.00	
	159		

[a] Letters at left of sources match with letter next to the error term used to compute *Fs*.
** $p < .05$.
**** $p < .001$.

Testing for perceptual reactivity. Testing for perceptual reactivity to the prepared room was conducted under two perceptual sets: casual perception and directed perception.

Casual perception. Immediately after the subject had "waited" 4 minutes in the prepared room, he was taken to another room and asked to report everything that he had observed during his "wait." He was given 3 minutes to report his observations while the experimenter checked off his responses on a mimeographed sheet listing the room contents.

Directed perception. Following the report of his casual observations, each subject was taken back to the prepared room and told that he now had 2 minutes in which to report his observations of the room he saw before him.

Design. A 2 × 2 × 2 factorial analysis of variance for one independent and two correlated variables was employed. The independent variable consisted of the two personality styles: expansive and restrictive; the two correlated measures consisted of number of figure and ground objects scored under casual and directed perception conditions. Forty subjects were used in this analysis; the remaining four were randomly removed from the restrictive group in order to obtain equal *Ns* for each personality style.

RESULTS

Table 2 presents results from the analysis of variance which bear on the three hypotheses previously formulated.

Hypothesis 1. The expectation that expansive people would note significantly more objects than restrictive people was not supported by the source of variance reflecting differences between the two styles ($F = < 1.00$, $df = 1/38$, $p > .05$).

TABLE 3

MEAN FIGURE AND GROUND OBJECTS SCORED FOR EXPANSIVE AND RESTRICTIVE STYLES

Object	Expansive	Restrictive
Figure	30.4	30.6
Ground	25.8	22.6

Hypothesis 2. The expectation that expansive people would note relatively more ground and less figure objects than restrictive people, for whom the opposite relationship was anticipated, was supported by the significant interaction between personality styles and kind of objects ($F = 4.92$, $df = 1/38$, $p < .05$) and by the direction of the differences between the means in Table 3. This table also suggests that the absence of a significant main effect between styles was due to the unanticipated similarity of figural object counts for people in both styles. It indicates that differences in style are most accentuated with regard to perception of ground objects.

Hypothesis 3. This hypothesis anticipated that people would manifest their styles independent of volitional changes in their attitude toward what they perceived. Two related sources may be adduced for this hypothesis. The first is based on the fact that casual and directed perception treatments were pooled and thus contributed to the significant interaction between styles and objects. The second source of indirect support is based on the nonsignificant first and second order interactions which include both the styles and perceptual attitude variables. Since both these sources of variance provide an opportunity for style to vary as a function of change in perceptual attitude, and both are nonsignificant, the null hypothesis that styles are stable independent of perceptual attitude cannot be abandoned.

Secondary findings. Secondary findings from Table 2 indicate that the distinction between figure and ground objects, favoring perception of figure objects, is supported ($F = 67.88$, $df = 1/38$, $p < .001$). Also found is a significant difference between perceptual attitudes ($F = 257.70$, $df = 1/38$, $p < .001$) indicating that people noted strikingly more objects when they were looking purposefully at them than when viewing them casually. Of greater interest was the unanticipated finding that perceptual attitudes interact significantly with kind of objects ($F = 6.46$, $df = 1/38$, $p < .05$). This interaction indicates that when perception is purposefully directed many more figural objects are noted than ground objects while when perception is casual, the number of figure and ground objects noted, although still favoring figural objects, are more nearly the same.

DISCUSSION

Present findings provide qualified support for the view that expansive and restrictive personality styles may be extended to include visual perception. The qualification exists because the first hypothesis that expansive people would perceive more objects within their immediate environment than restrictive people was not supported by a statistically significant difference. Nevertheless, support for the style generalization view exists because the second hypothesis was confirmed and there was evidence in favor of the third hypothesis. Two things require further consideration: first, the process whereby these personality styles might be generalized to visual perception; second, the unexpected finding that people of both styles perceived about the same number of figural objects.

With regard to the first, it is suggested that personality style may be generalized to visual perception via the motor behavior (eye movements) of the organism. Sweeping eye movements with brief fixations would be consistent with the expansive style while more limited eye movements with longer fixations would be consistent with the restrictive style. It is well established (Buswell, 1922; Tinker, 1936) that those people who manifest the broadest eye sweeps (saccadic movements) and shortest fixations are able to read more words per time interval than those who have more limited saccadic movements and longer fixations. If the reading of words in sentences is analogous to noting ground objects in a room, present findings are consistent with the view that these styles

may be reflected in visual perception by stylized eye movements.

The finding that both styles perceived about the same number of figural objects may also be explained in terms of eye movements. Figural objects are thought to have been too attractive, or, in terms of eye movements, to have induced shorter saccadic movements with longer fixations, limiting the number of figural objects noted, and thus partially obscuring the expected effects of style. Another study which reduced the valence of figural objects might find the greater overall environmental reactivity by expansive people only partially found in the present study.

SUMMARY

It was hypothesized that groups of expansive and restrictive people, selected on the basis of gesture, speech and ratings, would generalize their styles to visual perception. Partial support for style generalization was apparent in the finding that expansive people noted a greater number of background objects than restrictive people. Contrary to expectation was the failure to find that expansive people would perceive, independent of kind of object (figural or ground), a significantly greater number of objects than restrictive people.

REFERENCES

ALLPORT, G. W., & VERNON, P. E. *Studies in expressive movement*. New York: Macmillan, 1933.

BUSWELL, G. T. Fundamental reading habits: A study of their development. *Suppl. educ. Monogr.*, 1922, No. 21.

EISENBERG, P. Expressive movements related to feelings of dominance. *Arch. Psychol., N. Y.*, 1937, No. 211.

TINKER, M. A. Time taken by eye movements in reading. *J. genet. Psychol.*, 1936, **48**, 468–471.

(Received August 27, 1961)

Journal of Abnormal and Social Psychology
1962, Vol. 65, No. 5, 338–342

CASE REPORT

SOME EFFECTS OF EARLY SENSORY DEPRIVATION ON LATER BEHAVIOR:

THE QUONDAM HARD-OF-HEARING CHILD

NATHANIEL HERMAN EISEN

University of Illinois

Within the past few years there has been an increasing interest in the effects on adult behavior of temporary sensory deprivation during early stages of the developmental process. As an area of interest to psychologists, it stems primarily from Hebb's (1949) theoretical neurophysiological model of the function of the nervous system, and psychoanalytically oriented interpretations of the effects of mother-child separation on later mental health (e.g., Bowlby, 1951). Among other contributions to the picture have been Piaget's (e.g., 1954) conceptualizations of various aspects of the child's intellectual development, and Fuller's (e.g., Fuller & Thompson, 1960) work on the effects of early genetic-experiential interactions on the development of later behavior. Recently Hunt (1957, 1961) has made an extensive analysis of the observations, experiments, and theory in this area, and has developed a unifying point of view. One reason for the importance of this area for both psychological theory (e.g., Bruner, 1961) and application (e.g., Richmond & Lipton, 1961) is the possibility that such temporary sensory deprivations, if they occur at certain crucial stages of the developmental process, may prove to have irreversible effects.

Although there has been a rich interaction among observation, theory, and experiment in this area, for obvious reasons the experimental work has been done largely with infrahuman organisms, including rats, cats, dogs, and primates. When early sensory deprivation has been induced experimentally, it has to a great extent been either deprivation of vision or nonspecific social deprivation. Clinical observations of the effects of early sensory deprivation have either involved deprivation of vision or nonspecific deprivations under the rubric of "maternal deprivation." To the writer's knowledge, there have been only two studies reported of the effects of early temporary auditory deprivation on later behavior, and both of these were experimental studies of the behavior of rats; one was the classic study of Wolf (1943) and the other a recent replication of it by Gauron and Becker (1959).

The writer, a clinical psychologist, has over a number of years seen several children with adjustment problems which may be ascribed to temporary auditory deprivation during the first 2 or 3 years of life. Following is a typical case history of a "quondam hard-of-hearing" child.

James, a 7-year-old white boy, was referred for psychological evaluation because he was having adjustment problems in school. In the first grade he had been inattentive, had made a poor social adjustment, and had had learning difficulties. His teacher, concerned about his poor attention and his delayed reaction to questions and instructions, had considered the possibility that James was hard-of-hearing, and had referred him for a hearing examination. The otologist reported that James' hearing was well within normal limits. James' behavior showed no change, the process was repeated soon after he entered second grade on the basis of a "social promotion," and it was finally decided that the problem must be "psychological" or "emotional" in origin.

Psychological evaluation showed James to be capable of functioning at least within the average range. Nothing was obtained from the play interview, the psychological test data, or the history to suggest the possibility of brain damage, psychosis, or autism; James was able to relate well, he showed adequate reality testing for his age, and adequate performance in tasks of perception and coordination. Also eliminated as possible explanations of the reported behavior were malingering and such neurotic disorders as conversion reaction. Malingering, as a form of deviant behavior, must have some motivation such as the child's desire to express hostility towards authority figures, or it must provide some reward such as special attention, and no such motivation or rewards could be demonstrated in this case. Conversion reaction symptoms usually are associated with primary and secondary gains, and these could not be demonstrated with regard to James either.

James' behavior during the examination and in response to the test materials showed many of

the characteristics of a hard-of-hearing child. He had difficulty in localizing sounds, and at various times could not tell whether background noises originated in the testing room or in the street five floors below. Words out of context were not discriminated clearly, and instead of defining them he sometimes gave clang associations (e.g., donkey-duck; nitroglycerine-nightmare). His fund of general information and his ability to find words to express his thoughts were relatively impoverished, as compared with his overall ability as estimated from nonverbal tests. He had great difficulty in tasks requiring abstraction or conceptualization. He also seemed to have some difficulty in understanding what was expected of him in social situations, nor did he seem to understand why other people acted as they did in such situations.

James' history revealed that he had had middle ear infections intermittently during the first 3 years of his life. These apparently had not been painful, and had gone unnoticed for some time. However, the parents became concerned when James showed retarded speech development, and because he "did not pay attention." Medical care seemed to alleviate the source of the infection, and James has had no difficulty with his hearing since the age of 3. However, the failure to "pay attention" has persisted, and the parents have attributed it to malingering. They have been in turn overprotective and punitive in response to this aspect of James' behavior; the mother, especially, has been extremely inconsistent in dealing with James, since she herself had punitive parents and was very much rejected by her siblings for her "bad behavior" as a child.

Although the parents reported that James' speech development progressed in a normal and adequate manner, once the hearing difficulty was cleared up, both his teachers and the present writer noticed a peculiarity in his communication. There seemed to be a slight delay between stimulus and response, between question and answer, almost as if James were "thinking over" or "translating" speech into some more meaningful method of communication. It was this behavior that led his teachers to request a hearing examination.

Further inquiry revealed that James' inattentiveness seemed to increase when he was in a stressful situation, such as when he was called upon to recite in class, or when he was reprimanded at home. A similar pattern appeared in response to the psychological tests; the more stressful or complex the task, the more likely was James to ask to have the question repeated, or to indicate in other ways that he had not heard what had been said to him.

The total picture reminded the writer of the behavior reported by Wolf (1943) for his experimental rats; animals temporarily deprived of their hearing during their nursing period and who then had their hearing restored seemed to adjust adequately in later life with full development of their sensory resources, until their security was threatened by their being placed in a stressful competitive situation. At this point the mature responses seemed to disintegrate and the rats acted as if they were still unable to hear. On the basis of this similarity, the writer has attempted to explain the problems of James and of children like him (the "quondam hard-of-hearing") using concepts developed out of the research and clinical observations in the area of sensory deprivation.

The language in which these concepts have been expressed has varied from writer to writer, and has been drawn from a variety of theoretical sources currently popular in psychology—perceptual theory, learning theory, information theory. As succinct a statement as any is that of Bruner (1961):

> In general, an impoverished environment, one with diminished heterogeneity and a reduced set of opportunities for manipulation and discrimination, produces an adult organism with reduced abilities to discriminate, with stunted strategies for coping with roundabout solutions, with less taste for exploratory behavior, and with a notably reduced tendency to draw inferences that serve to cement the disparate events of its environment. . . . To these evidences of reduced capacities, which may indeed be irreversible, although there has not yet been a full-scale attempt to provide adequate therapy toward overcoming these deficits, we should add the fact that there seem to be critical periods operative. Unless certain forms of stimulation-cum-learning take place before a certain point in (an organism's) life, there appear to be certain very intractable changes (p. 199).

The child who, like James, has had hearing difficulties early in life, during those crucial stages in which one learns to listen to the speech of others, to pay attention, to respond with speech and to develop relationships with others, may never learn to do these things in a normal or adequate manner. He fails to make fine discriminations among sounds, or to localize them. Lacking basic attitudes with regard to interpersonal communication, he fails to acquire the vocabulary and the general information acquired by the normal child who listens, pays attention, and attempts speech. As maturation proceeds, he does not have the necessary foundation out of which

develops the use of implicit speech or thought. This in turn leads to inadequate development of higher order processes of abstraction and conceptualization. In school he may have difficulty learning to read.[1]

Once the hearing difficulty disappears the child may learn to use speech in a manner that seems normal on superficial examination. However, the reported delay between stimulus and response, the impression James gives of "not listening or paying attention," suggests that this may not be the case. Speech for children like James may not become the easy basis for interpersonal relationships that it is for most people. If this is the case, we would expect the more complex and integrated aspects of behavior dependent on normal relationships—social behavior, emotional responsiveness—to develop in atypical fashion. This is what seems to have happened to James.

The relationship between early temporary auditory deprivation and later behavioral or adjustment deviations does not seem to be a simple or direct one, however. According to Wolf (1943), his rats showed the effects of early deprivation primarily under the stress of competition; under conditions of isolated performance, the two groups performed alike. A similar situation seems to exist for the quondam hard-of-hearing child; competition seems to exacerbate if not to elicit the deviant behavior. (School might be considered a competitive situation for any child; in James' case, although his IQ was 106, this placed him second from the bottom of a class in which the average IQ was in the 120s.) It is likely, therefore, that hard-of-hearing-like behavior in children who have been subject to early temporary auditory deprivation is a function not only of this early deprivation, but of the child's ability to tolerate competition, stress, or frustration of various kinds. We are dealing, it seems, with the effects of the interaction of at least two sets of variables.

But the child's ability to withstand stress is itself the resultant of several factors in his life history. One of these would be the specific reactions of his parents to his early hearing difficulties—their tendency to ignore it, to become overly concerned about it, to take it in stride, etc.; that is, whether they become overprotective, overly rejecting, or showed the "normal" or "proper" amount of concern. Another factor influencing the child's ability to withstand stress would be the parents' general child rearing techniques, irrespective of the child's hearing difficulties (insofar as the interactive effects could be partialed out). All of the quondam hard-of-hearing children seen by the writer have come from atypical or abnormal familial backgrounds; broken homes, extreme socioeconomic impoverishment, parents who were themselves emotionally disturbed in some way. From the histories of these children, he has the impression that these children might have developed in abnormal fashion even if there had been no sensory deprivation; the fact that they were deprived during a crucial stage of development merely provided the avenue into which the abnormality was channeled.

In other words, before the quondam hard-of-hearing child shows behavior deviant enough to come to the attention of the psychologist, there has to have occurred the interaction of a number of variables:

A. Predisposing factors

1. Biological

a. Infection or other disease leading to hearing loss during crucial stages of preverbal and early verbal development. This clears up at a later stage.

2. Psychological

a. A family background leading to faulty personality development—at least a lowered tolerance for stress or competition.

b. Familial reaction to the hearing difficulty so as to intensify the lowered stress tolerance.

B. Precipitating factors

In most instances, the stresses and strains of adjustment to school, peers, and other social factors that the child encounters as he leaves the more sheltered environment of the home.

All of the above factors interact within the matrix of a society in which there are certain demands and expectations with regard to the use of speech, the "paying of attention," etc. All other factors being equal, we would not expect the quondam hard-of-hearing child to come to the psychologist, or to be a matter of concern, unless these social expectations existed.

It follows from this analysis that there may be an inverse relationship between the degree of early temporary auditory deprivation (either in severity or duration) and the other factors. This seems to be in accord with the observations of this writer.

As was noted above, interest in the effects of sensory deprivation derives in part from the possibility that, if the deprivation occurs before what may be a critical period in the life of the

[1] J. J. O'Neill, personal communication, 1961.

organism, the effects may prove to be irreversible. However, as Bruner (1961) has pointed out, there has not yet been a full scale attempt to provide adequate therapy toward overcoming the effects of such deprivation. The clinical evidence, as summarized by Hebb (1949), Hunt (1957, 1961), and others is equivocal, since the possibly relevant variables are confounded to an unknown degree. (The effects of therapy may prove to vary as a function of the position of the organism on the phylogenetic scale and the sensory modality in which deprivation occurs, as well as of the more obvious variables such as age of onset of deprivation, duration, and techniques of therapy used.) Gauron and Becker (1959) found that their rats learned to compete better with practice, and that "by the last four days of the second competition the advantage of the nondeprived animal is hardly noticeable" (p. 692). However, the experimental design did not permit a clear inference as to the reason for the apparent recovery of function; age, practice, and dominance effects were confounded.

While no systematic attempt was made to provide therapy for the few quondam hard-of-hearing children seen by the writer, two of them were able to obtain therapy. One of the children was seen for a period of several months by the writer, and reports from both his home and from the school indicated changes in behavior. However, these changes may have been due to a reduction in the external precipitating stresses, rather than to any modification of the more basic personality predispositions for the symptomatic behavior.

James was the other child provided with therapy; he was seen by a school counselor for several months, while the mother was seen by the writer for about five sessions. In addition, a deliberate attempt was made to reduce the external sources of stress by having him placed in a less competitive classroom situation, and by explaining to his principal and to his new teacher the sources of his difficulties, as hypothesized above. As in the previous instance, reports from both home and school indicated improvement in behavior.

The writer, with the idea of doing a follow-up study of the effects of therapy on James, contacted his parents just about 2 years after he had first been seen. The mother's first words on the phone were, "I was just thinking of calling you; we have been concerned about James again." He was having difficulty both at home and in school—attention problems, learning problems, social adjustment problems. He was therefore seen again for a complete diagnostic evaluation.

The diagnostic picture was much different this time. While James' recall vocabulary was still relatively impaired, his fund of general information was at a level expected of one of his overall ability. His ability to conceptualize had also improved. He no longer showed the lack of understanding of what to do or say in social situations, and in fact had developed a great deal of poise and even charm. However, further investigation suggested that this was a mere façade, and that beneath his bright-eyed smile lay intense feelings that the world was a very threatening place. James' defense against the world was emotional withdrawal covered up by a façade of what seemed to be charm and poise. The strain of maintaining this façade made itself apparent in the very real adjustment difficulties he was experiencing. In this James very much resembled his mother, who had made a similar adjustment to life as a result of her own childhood experiences. Even his bright-eyed smile was very much like his mother's.

The specific details of this diagnostic picture are less important than the fact that James no longer resembled the other quondam hard-of-hearing children seen by the writer, nor did he present the same picture he had presented 2 years before. While some of the characteristics which had led the writer to label James a quondam hard-of-hearing child were still present, they were far overshadowed by more "typical" neurotic manifestations. Had the writer not known the boy's history, and had he not evaluated the boy earlier, he would have overlooked those characteristics that have here been attributed to early temporary sensory deprivation, or he would have attributed them to severe emotional stress.

James' history does not answer the question as to the conditions under which the effects of early temporary sensory deprivation might be reversible. But it does suggest that some of these effects might be modifiable under certain as yet unknown conditions.

As was pointed out above, the effects of early temporary sensory deprivation are of both practical and theoretical importance. In dealing with deviant children, clinicians have a tendency to overlook those constitutional or biological characteristics that may serve as predisposing factors towards deviant behavior. The writer has seen several children, exemplified by James as described above, whose deviant behavior seems due

in part at least to the fact that they suffered early temporary auditory deprivation. While the deprivation in these cases does not explain by itself the fact that deviation occurred, it serves as a predisposing factor and may explain the direction of the deviation. However, further research is needed before the description of the quondam hard-of-hearing child can be accepted as a bona fide syndrome.

The reversibility of the effects of early temporary sensory deprivation is of great theoretical as well as practical interest. As is obvious, experimental work in this area is possible only with infrahuman organisms; with such organisms it is not possible to learn directly of the effects of auditory deprivation on speech and on the higher thought processes. If the quondam hard-of-hearing child presents a bona fide syndrome, we have here an "experiment of nature" and it may be possible to use these children as subjects of research into a number of areas, including the effects of variations in therapeutic and educational techniques in modifying behavior and personality patterns established early in life.

REFERENCES

BOWLBY, J. Maternal care and mental health. *Bull. WHO*, 1951, **3**, 355–533.

BRUNER, J. S. The cognitive consequences of early sensory deprivation. In P. Solomon et al. (Eds.), *Sensory deprivation: A symposium held at Harvard Medical School*. Cambridge: Harvard Univer. Press, 1961. Pp. 195–207.

FULLER, J. L., & THOMPSON, W. R. *Behavior genetics*. New York: Wiley, 1960.

GAURON, E. F., & BECKER, W. C. The effects of early sensory deprivation on adult rat behavior under competition stress: An attempt at replication of a study by Alexander Wolf. *J. comp. physiol. Psychol.*, 1959, **52**, 689–693.

HEBB, D. O. *The organization of behavior*. New York: Wiley, 1949.

HUNT, J. MCV. Investigations of the effects of early experience in sub-human animals. Unpublished manuscript, University of Illinois Library, 1957.

HUNT, J. MCV. *Intelligence and experience*. New York: Ronald, 1961.

PIAGET, J. *The construction of reality in the child*. New York: Basic Books, 1954.

RICHMOND, J. B., & LIPTON, E. L. Studies on mental health of children with specific implications for pediatricians. In G. Caplan (Ed.), *Prevention of mental disorders in children*. New York: Basic Books, 1961. Pp. 95–121.

WOLF, A. The dynamics of selective inhibition of specific functions in neurosis. In S. S. Tomkins (Ed.), *Contemporary psychopathology*. Cambridge: Harvard Univer. Press, 1943. Pp. 398–413.

(Received July 16, 1961)

Journal of Abnormal and Social Psychology
1962, Vol. 65, No. 5, 343–347

CRITIQUE AND NOTES

IDENTIFICATION AND EMPATHY IN RELATION TO SELF-DISSATISFACTION AND ADJUSTMENT [1]

DORIS H. LOCKWOOD AND BERNARD GUERNEY, JR.

Rutgers University

A parent's empathy with his child, the child's self-concept and its development, and the child's identification with his parents are variables that have been widely used in the investigation of a child's personality development and emotional adjustment. Attempts to relate parental empathy to children's adjustment have not been particularly successful (e.g., Baker, 1956; Langford & Alm, 1954). On the other hand, investigations relating the self and its discrepancy with the ideal-self (self-dissatisfaction) to other measures of emotional adjustment, using adult subjects, have been more successful (Hanlon, Hofstaetter, & O'Connor, 1954; Lepine & Chodorkoff, 1955); and studies of parent-child identification as it relates to the child's adjustment (Payne & Mussen, 1956) have also been encouraging. But there have been few attempts to relate self-dissatisfaction to parental empathy and identification, or to investigate the self–ideal-self discrepancy as a variable in the adjustment of young subjects. The present study was undertaken in order to explore the interrelations of empathy, identification, and self-dissatisfaction; to employ subjects who were young enough to reflect active parent-child interactions, but old enough to use self-report instruments; and to try to draw upon some of the more recent theories and concepts regarding empathy and self-report techniques. The following hypotheses were proposed:

1. The greater the empathy of the same-sexed parent with his child, the better will be the child's total adjustment, emotional adjustment, home adjustment, social adjustment, and lack of self-dissatisfaction.

2. The stronger the child's identification with the same-sexed parent, the better will be the child's total adjustment, emotional adjustment, home adjustment, social adjustment, and lack of self-dissatisfaction.

3. The greater the parent's empathy for his child, the more the child will identify with that parent.

4. The more dissatisfied a child is with himself, the poorer will be his total adjustment, emotional adjustment, home adjustment, and social adjustment.

METHOD

Materials

Cronbach (1950, 1955) and others (Bass, 1956; Edwards, 1957; Fricke, 1956) have recently called attention to some serious limitations of self-report instruments and of objective personality and adjustment inventories. These authors point out that variables such as empathy, identification, and self-dissatisfaction, when measured by self-report and objective tests, are often confounded with response sets such as acquiescence, projection, and social desirability. They suggest various methods of controlling for these response sets, and several of these methods were, in fact, applied to the present data. But, because of the broad and exploratory nature of the present investigation, and because of the dearth of well-controlled tests suited to the pertinent variables, the use of a self-report instrument and an objective personality test seemed justified.

The Interpersonal Check List (ICL), developed by Rolfe LaForge and Robert Suczek (Leary, 1957), was chosen to measure empathy, identification, and self-dissatisfaction. The individual checks an item (e.g., "well thought of," "self-confident," etc.) if he feels it is descriptive of the person he is considering, and leaves it blank if it is not descriptive. Normally, ICL scores yield a measure of 16 interpersonal variables, but for present purposes, an item comparison was used instead (see below).

The Student Form of the Bell Adjustment Inventory provided measures of home, social, and emotional adjustment. The Total Adjustment score is based on all the scales of the inventory, but note that Health Adjustment, a component of Total Adjustment, was not used separately in this investigation. The subscales of the Bell are not highly intercorrelated; relevant correlations (as reported in the manual) are: home and social, .04; home and emotional, .38; social and emotional, .47.

Subjects

The subjects were 40 high school students, 20 boys and 20 girls, all members of a "home room" group

[1] This paper is based on a thesis submitted by Doris H. Lockwood in partial fulfillment of the requirements for the MA degree at Rutgers. Thanks are extended to Benjamin Rosner and Oscar Buros for their suggestions regarding statistical procedures, and to Nelson Hanawalt, Jean Burton, and Winfield Scott for their interest and assistance in this study.

which had volunteered for the study. The boys' ages ranged from 14 to 16 years, with a mean of 14.6 years; the girls' ages ranged from 13 to 16 years, with a mean of 14.4 years. The group was told that it was participating in a study of the way high school students feel about themselves. In the latter half of the study, for reasons given below, the subjects were reduced to 28—15 females and 13 males.

A Henmon-Nelson Test of Mental Ability, administered 8 months prior to the study, revealed the boys' IQs to range from 101 to 144, with a mean of 116.6; the girls' IQs ranged from 85 to 128, with a mean of 106.4. Fathers' occupations, classified according to the United States Department of Labor (1949), were wide in range, but weighted toward the highest level. There were 13 Professional and Managerial; 3 Clerical and Sales; 1 Service Occupations; 2 Agriculture, Fishery, and Forestry; 8 Skilled Occupations; 4 Semiskilled Occupations; and 6 Unskilled. Three fathers remained unclassified because of retirement or absence from home.

Procedure

Two separate sessions were required to obtain the data. At the first session, the ICL was administered, and the subjects were asked to fill it in as follows: as they saw themselves at the present time: "What am I like"; as they wished they were, their ideal: "What I would like to be." After the test was begun, all questions were referred to the directions on the front of the test booklet. Difficult words were defined on request, using a dictionary definition. At a second session of the same class, exactly 24 hours later, all subjects were asked to fill out the Bell Adjustment Inventory. Test instructions were followed as outlined in the manual for the Student Form (Bell, 1934).

Several weeks later, an ICL form, coded for identification purposes, was sent to the parent of the same sex as the child. The parent was asked to fill it in by describing himself and as he thought his child would describe himself. After 2 months, eight parent forms were still outstanding, and the children of these parents were dropped from this half of the study, as were four children living with the parent of the opposite sex because of death or divorce.

Measurement of the Variables

The data were defined and measured in the following manner:

Identification (in the sense of becoming identical through affiliation with a person) was operationally defined as follows: Each item on the child's ICL self-description ("What I am like.") was compared with the corresponding item on the parent's ICL self-description. If the two items were alike—that is, either both checked or both left blank—the identification score was marked 1+. The total identification score was the number of plus marks. The investigators felt that the actual similarity between a child's and parent's self-ratings, whether the child was conscious of the similarity or not, was a better measure of identification than only the child's judgment of such similarity, which would probably be more contaminated by such variables as social desirability and projection.

Self-dissatisfaction was operationally defined as follows: Each item for the child's self-description was compared with each item of the child's ideal-self-description ("What I would like to be."). If both were the same, either checked or left blank, no score was tallied. If they were different, the self-dissatisfaction score received 1+. The total score was the number of plus marks.

Empathy (in the sense of comprehension of the state of mind of another person) is often measured by having one individual predict how another individual will describe himself on some test or check list; the number of correct predictions reflects empathy. Hastorf and Bender (1952) point out that some individuals tend more than others to project their own characteristics into their predictions. Such projection would then contaminate empathy. But, one can obtain a refined empathy score by, in effect, subtracting the correctly predicted items which are the same as the predictor's self-description score (the projection score) from the total number of correct predictions (raw empathy score). Hastorf and Bender further point out that if the predictor and the predictee have similar personalities, the use of this method might create spuriously high projection scores and therefore spuriously low empathy scores. In the present study, it would also make for a spuriously low or negative correlation between identification and empathy. In order to minimize the possibility of inflating empathy scores with projection, and to minimize penalizing parents whose personalities were genuinely similar to their children's, the following procedure was adopted: first, to count all correct predictions; second, to subtract from this total not *all* items which were the same for the parent's self-description and predicted scores, but only those which were also *correct* predictions; and third, to make the final empathy score a *percentage* of the correct answers remaining after the discarding process, rather than the absolute number.

RESULTS

Means and SDs

A summary of means and *SD*s is shown in Table 1. "Assumed similarity" is a post hoc variable explained below. For identification and self-dissatisfaction, male *SD*s were almost three times as large as female *SD*s. Note that the empathy means are only about 25 for both groups, a figure which is not surprising, since "projection" eliminated some correct predictions. For both boys and girls, the means of the Bell scales in this study fell within Bell's "average adjustment" category.

TABLE 1

MEANS AND STANDARD DEVIATIONS OF SELF-DISSATISFACTION, IDENTIFICATION, PARENTAL EMPATHY, ASSUMED SIMILARITY, AND BELL ADJUSTMENT INVENTORY SCALES

Variable	Males ($N = 13$)		Females ($N = 15$)	
	M	*SD*	*M*	*SD*
Self-dissatisfaction	29.3	22.9	32.1	8.6
Identification	85.8	14.6	87.5	5.6
Empathy	26.1	7.1	24.8	7.8
Assumed similarity	94.0	17.4	100.7	10.7
Total adjustment	35.2	16.7	48.1	15.6
Emotional adjustment	9.0	7.8	15.1	6.9
Home adjustment	7.6	6.9	11.0	7.4
Social adjustment	12.5	4.8	15.1	5.6

Note.—The lower the score on the Adjustment Inventory, the better the adjustment.

Tests of Hypotheses

The Pearson product-moment coefficient of correlation, and two-tailed tests of significance at the 5% level of confidence were used throughout. These correlations are summarized in Table 2.

To test the first hypothesis, empathy was correlated with the Total, Emotional, Home, and Social Adjustment scales of the Bell Adjustment Inventory, and with ICL self-dissatisfaction scores. None of these correlations was significant; the first hypothesis was not supported.

To test the second hypothesis, identification was correlated with the Total, Emotional, Home, and Social Adjustment scales of the Bell inventory, and with ICL self-dissatisfaction scores. There was a difference between male and female results: male correlations (with the exception of social adjustment) were statistically significant, while female correlations were insignificant. Boys who identified strongly with their fathers were better adjusted, as measured by the Total, Emotional, and Home Adjustment scales of the Bell.

The correlations of identification with ICL self-dissatisfaction scores revealed similar trends with respect to sex differences: there were significant correlations for males (again excepting social adjustment) and insignificant correlations for females. Boys who identified strongly with their fathers had less discrepancy between their self-concept and their ideal-self-concept (less self-dissatisfaction). With the exception of social adjustment, the second hypothesis was supported for boys, but not for girls.

The third hypothesis, that parental empathy is associated with greater child identification, was tested by correlating empathy with identification. These correlations were not significant, and the third hypothesis was not supported.

To test the fourth hypothesis, ICL self-dissatisfaction scores were correlated with the Bell scales. Sex differences were again apparent. The boys revealed positive, significant correlations for total, emotional, and home adjustment; that is, the greater their self-dissatisfaction, the poorer their total, emotional, and home adjustment. (The correlation with social adjustment again failed to reach significance.) The girls' correlations were not significant. Thus, the fourth hypothesis was partially supported with respect to boys.

In order to clarify the insignificant results for empathy, another variable was investigated: assumed similarity. The score for this variable denoted the similarity between the parent's self-description and his prediction for his child's self-description, whether this prediction was correct or not. Fathers' assumed similarity scores were correlated with their sons' total, emotional, and home adjustment scores, and with identification scores. The correlation with emotional adjustment was not significant ($r = -.50$). Total ad-

TABLE 2

CORRELATIONS BETWEEN IDENTIFICATION, SELF-DISSATISFACTION, AND BELL ADJUSTMENT VARIABLES

Variable	*N*	Identification	Self-Dissatisfaction	Total Adjustment	Emotional Adjustment	Home Adjustment	Social Adjustment
Empathy							
Girls	15	−.28	.12	−.29	−.30	−.48	−.01
Boys	13	−.15	−.23	.54	.31	.48	−.06
Identification							
Girls	15	—	−.31	.18	.32	.06	.31
Boys	13	—	−.69**	−.65*	−.66*	−.65*	−.02
Self-dissatisfaction							
Girls	20	—	—	.09	.03	−.16	−.13
Boys	20	—	—	.49*	.47*	.47*	−.17

Note.—Lower scores indicate better adjustment.
* Significant at the .05 level, two-tailed test.
** Significant at the .01 level, two-tailed test.

justment ($r = -.65$) and identification ($r = .67$) were significantly correlated with assumed similarity at the 5% level. Home adjustment ($r = -.72$) was significant at the 1% level of confidence. The more a father assumed similarity to his son, the better was the son's total and home adjustment. And, the more the father assumed similarity to his son, the stronger the son's identification with his father.

Response Set to Check Items

It was felt that a response set to check items on the ICL might have been causing spuriously high correlations between some of the variables. A superficial tendency to "acquiesce" to items on the adjustment scales and to items on the ICL might spuriously increase the correlation between the discrepancy scores and adjustment scores. Therefore, partial correlations were run on measures yielding significant results in order to hold constant the number of ICL items checked. With this procedure, male correlations for identification with total, emotional, and home adjustment were still significantly high (−.67, −.65, −.63, respectively). Male correlations for fathers' assumed similarity with total and home adjustment also remained significantly high (−.74 and −.77, respectively). It was concluded that a response set for checking items was not causing these significant relationships.

DISCUSSION

The results indicated that empathy, as operationally defined in this experiment, was not correlated with either male or female adjustment as measured by the Bell Adjustment Inventory, or with self-dissatisfaction as measured by the ICL. Because of these findings, which were contrary to widely held views, it was felt that the procedures used in eliminating "projection" were not appropriate. Another familiar hypothesis was applied: that seeing others, whether accurately or not, as similar to oneself (assumed similarity) may positively influence the interpersonal relationship. The highly significant male correlations of assumed similarity with both measures of adjustment—the Bell scales and the self-dissatisfaction score on the ICL—suggest that this variable (as operationally defined here) may be relevant in an analysis of parent-child relationships. Cronbach (Tagiuri & Petrullo, 1958) would probably suggest that the variables entering into such correlations must be analyzed by better controlled procedures than those employed in the present study. All that can be said is that from the present data it appears that some unknown expressed personality trait of the father seems to facilitate his son's adjustment and his son's identification with him. That fathers' assumed similarity scores correlated significantly with their sons' adjustment (as measured by a separate test) despite the use of "superficial" instruments of measurement, indicates the need for further investigations along these lines.

Another striking feature of the results is that although female correlations were not significant for many of the variables, male correlations were. There are a number of possible reasons for this. The small sample used in this study may have contributed towards sampling errors. On some of the variables, the girls' standard deviations were much smaller than the boys', and showed, as a possible consequence of sampling errors, less variability and individuality in responses. On the other hand, this may have reflected true sex differences. Further investigation is needed to clarify this problem. Another possibility is that daughter-father relationships at this age are more pertinent to the variables in question than daughter-mother relationships. Sopchak (1952), for example, while investigating college students' identification with their parents, also failed to get results for mother-daughter identification.

The significant male correlations between boys' ICL self-dissatisfaction scores and their total and emotional adjustment on the Bell, suggests that the ICL, used in this way, might have some validity as a measure of adjustment. More research, using other measures of adjustment and a larger sample, is indicated, particularly since female results were insignificant.

SUMMARY

The purpose of the present study was to investigate the relationship among the variables empathy, identification, self-dissatisfaction (self–ideal-self discrepancy), and adjustment. Two instruments, the Interpersonal Check List and the Bell Adjustment Inventory, were used. The subjects were 40 high school students (20 males and 20 females). Of these 40 children, 13 boys and their fathers and 15 girls and their mothers participated in the second part of the study. The ICL and the Bell inventory were administered to the students in two separate sessions. Later, the parent of the same sex as his child took the ICL, filling it in twice: first for himself, and then to predict his child's self-description.

For several of the variables, statistically significant results were revealed for males, but not for females. The more a boy identified with his father, the better was the son's total, emotional, and home adjustment, and the less his self-dis-

satisfaction. When all possibility of projection was eliminated, empathy, as operationally defined, revealed a mean score of only 25 for both fathers and mothers. A post hoc variable, assumed similarity of the parent to his child, correlated significantly with several of the variables for males. The more a father assumed similarity to his son, the better was his son's total and home adjustment, and the greater was the son's identification with him. It was suggested that parental assumed similarity might have psychological relevance to the child's adjustment. Male ICL self-dissatisfaction scores correlated significantly with several of the Bell scales, indicating its possible use as a measure of adjustment.

REFERENCES

BAKER, L. S. The relation of maternal understanding of the child and attitudes toward the child to adjustment of the child. Unpublished doctoral dissertation, New York University, 1956.

BASS, B. M. Development and evaluation of a scale for measuring social acquiescence. *J. abnorm. soc. Psychol.*, 1956, **53**, 296–299.

BELL, H. M. *Manual for the Adjustment Inventory, Student Form*. Stanford, Calif.: Stanford Univer. Press, 1934.

CRONBACH, L. Response set and test design. *Educ. psychol. Measmt.*, 1950, **10**, 3–31.

CRONBACH, L. Processes affecting scores on "understanding of others" and "assumed similarity." *Psychol. Bull.*, 1955, **52**, 177–193.

EDWARDS, A. L. *The social desirability variable in personality assessment and research*. New York: Dryden, 1957.

FRICKE, B. G. Response set as suppressor variable in the OAIS and MMPI. *J. consult. Psychol.*, 1956, **20**, 161–169.

HANLON, T. E., HOFSTAETTER, P. R., & O'CONNOR, J. P. Congruence of self and ideal-self in relation to personality adjustment. *J. consult. Psychol.*, 1954, **18**, 215–218.

HASTORF, A. H., & BENDER, T. E. A caution respecting the measurement of empathic ability. *J. abnorm. soc. Psychol.*, 1952, **47**, 574–576.

LANGFORD, LOUISE M., & ALM, O. W. A comparison of parent judgments and child feelings concerning self-adjustment and social adjustment of 12-year-old children. *J. genet. Psychol.*, 1954, **85**, 39–46.

LEARY, T. *The interpersonal diagnosis of personality*. New York: Ronald, 1957.

LEPINE, L. T., & CHODORKOFF, B. Goal setting behavior, expressed feelings of adequacy, and the correspondence between perceived and ideal-self. *J. clin. Psychol.*, 1955, **11**, 395–397.

PAYNE, D. E., & MUSSEN, P. H. Parent-child relations and father identification among adolescent boys. *J. abnorm. soc. Psychol.*, 1956, **52**, 358–362.

SOPCHAK, A. L. Parental identification and "tendency toward disorders" as measured by the MMPI. *J. abnorm. soc. Psychol.*, 1952, **47**, 159–165.

TAGIURI, R., & PETRULLO, L. *Person perception and interpersonal behavior*. Stanford, Calif.: Stanford Univer. Press, 1958.

UNITED STATES DEPARTMENT OF LABOR, Bureau of Employment Security. *The dictionary of occupational titles*. (2nd ed.) Vol. 1. *Definitions of titles*. Washington, D. C.: United States Government Printing Office, March 1949.

(Received July 17, 1961)

Journal of Abnormal and Social Psychology
1962, Vol. 65, No. 5, 347–351

THE RELATIONSHIP BETWEEN RESPONSE STYLE AND PERSONALITY VARIABLES:

II. THE PREDICTION OF INDEPENDENT CONFORMITY BEHAVIOR [1]

RICHARD K. McGEE

Moccasin Bend Psychiatric Hospital, Chattanooga, Tennessee

Research on the personal modes or styles with which individuals respond to test items has suggested that certain response tendencies may be useful predictors of underlying personality traits. It has recently been argued (Gray & Crisp, 1961; McGee, 1962a) that the use of the response set component of a test score for this purpose requires that the response tendency be a *general* one, not confined to a single pool of items. However studies by these investigators have shown that the evidence for a "pure" response set should be regarded with suspicion, and interpreted with caution.

[1] This study was completed while the author was at Vanderbilt University. Part of the investigation was submitted to the faculty in partial fulfillment of the requirements for the degree of Doctor of Philosophy. The dissertation project was directed by Samuel S. Komorita, who, along with David Marlowe, made helpful suggestions for the presentation of this report.

Too frequently data which support the contention that response sets are related to personality variables have contained a methodological bias inherent in the study design (McGee, 1962b). It is uncommon for investigators to discard the more easily administered psychometric scales in favor of independent behavioral tasks to measure the personality traits. Thus test taking behavior has been correlated with test taking behavior. Recent studies have shown that the set to make socially desirable responses may be conceptualized as a need for social approval (Crowne & Marlowe, 1960). Social Desirability scores have been shown to predict: conformity in a modified Asch situation (Strickland & Crowne, in press); the favorability of attitudes expressed after a boring task (Marlowe & Crowne, 1961); and the conditioning of verbal behavior (Crowne & Strickland, 1961; Marlowe, 1962). A thorough review of research on the agreeing response tendency (acquiescence) revealed only one major study (Jackson, 1959) where an independent behavioral task had been used to measure the dependent variable. This study related acquiescent responding to cognitive functioning, and indirectly suggested a connection between acquiescence and social behavior (Jackson, 1958). As a result, the hypothesis that acquiescers are socially oriented, extroverted, expressive conformers (Bass, 1956; Couch & Keniston, 1960; Gage, Leavitt, & Stone, 1957) remains to be examined.

Accordingly, the present study was designed to determine whether or not individuals who manifest a tendency to choose acquiescent options in response to test items would tend to display socially oriented *behavior*. It was hypothesized that: the acquiescent individual will yield or shift his own position (i.e., conform) under the influence of social pressure more than the nonacquiescer; and in a situation which demands social acceptance, the acquiescent individual will withhold negative feelings and profess feelings and attitudes consistent with winning approval more than the nonacquiescer.

METHOD

Seven measures of acquiescence and two independent behavioral tasks[2] were administered to 104 undergraduate students in an introductory psychology class. The subjects consisted of 60 males and 44 females ranging in age from 17 to 22 years.

Acquiescence measures. Five of the acquiescence measures were used in a previous study and have been described elsewhere (McGee, 1962b). The remaining two measures were the Reversed F Scale (Christie, Havel, & Seidenberg, 1958), and an item-by-item comparison of responses made to the 27 pairs of items common to the original F Scale (Adorno, Frenkel-Brunswik, Levinson, & Sanford, 1950) and the Christie et al. Reversed form. The subjects responded to each item by checking a seven-point agree-disagree scale. Scores were derived by summing the scale values over all items in a test, the higher scores indicating greater agreement tendency.

Behavioral measures. To test the first hypothesis regarding acquiescence and conformity, a modification of the Sherif (1935) autokinetic experiment was employed.

Apparatus. The stimulus light was produced by a 115-volt, 7-watt lamp glowing through an aperture approximately .5 millimeter in diameter. The stimulus was presented in a totally dark room and was located 56 inches above the floor, and 14 feet in front of the subjects. A multiple-cam timer operated the equipment such that a "ready" buzzer sounded 1 second prior to the activation of the lamp which glowed for 5 seconds during each 10-second cycle.

Procedure. Each subject was paired with a "stooge" of the same sex who was introduced to him as a graduate student. During Session I of the experiment, the subject and stooge were given 40 presentations of the stimulus and instructed to estimate the distance (in inches) they perceived the light to move during each exposure. These estimates were recorded privately on small pads of paper. This was followed by a 5-minute rest period during which the experimenter quickly computed the mean and standard deviation of the subject's last 20 judgments. These data were then secretly communicated to the stooge prior to Session II. During Session II 40 trials were given, however this time the judgments were reported to the experimenter verbally over a two-way intercommunication system. On half of these trials (randomly determined) the subject gave his judgment first, and on the other half the subject responded after hearing the stooge's judgment. The stooge varied his judgments around a value equal to the mean plus two standard deviations of the subject's Session I judgments. To measure the amount of shift or conformity, the mean of the subject's judgments (on the critical trials where he responded last) was computed, and compared with his Session I mean. The difference between these means was then divided by the amount of shift attempted by the stooge. Thus, the shift score for each subject is a ratio, based on his own initial performance, and shows the extent to which he varied his performance after the social influence was introduced.

The Spool Packing Task (Festinger & Carlsmith, 1959) was used to test the second hypothesis regarding the reporting of favorable attitudes to gain social approval. The procedure and instructions used in this study were the same as used by Marlowe and Crowne (1961) except that subjects packed and unpacked the spools for 30 minutes instead of 25 minutes. However, the reaction questionnaire was

[2] Copies of each of the acquiescence measures and forms for collecting the behavioral data may be obtained without cost by writing to the author.

TABLE 1

INTERCORRELATION MATRIX[a]

Variable	1	2	3	4	5	6	7	8	9	10	11
1. Social Acquiescence Scale	—	.12	.43	−.09	.35	.34	.34	.01	.00	.23	−.29
2. Agreement Response Scale	—	—	.15	−.14	.05	.05	.20	−.14	−.05	−.29	.22
3. California F Scale	—	—	—	−.24	.75	.27	.13	−.07	.09	.34	−.32
4. Reversed F Scale	—	—	—	—	.43	−.02	.12	.15	−.19	−.01	.08
5. Original-Reversed F	—	—	—	—	—	.24	.19	.04	−.06	.23	−.20
6. German Words	—	—	—	—	—	—	.16	.03	.10	.21	−.21
7. ESP Technique	—	—	—	—	—	—	—	−.01	−.14	.09	−.04
8. Spool Packing Task	—	—	—	—	—	—	—	—	−.10	.24	−.20
9. Autokinetic Conformity	—	—	—	—	—	—	—	—	—	.02	−.03
10. Marlowe-Crowne *SD* scale	—	—	—	—	—	—	—	—	—	—	−.37
11. Independence of Judgment	—	—	—	—	—	—	—	—	—	—	—
Correlation with Sex (r_{pb})[b]	.31	−.12	.24	.00	.18	.26	.16	−.01	.24	.50	−.35
Reliability	.83	.65	.73	.53	.77	.69	.30	.72	.99	—	—

[a] For correlations involving Variables 10 and 11, $N = 83$: $r = .22$, $p = .05$ and $r = .28$, $p = .01$. For all other test variables, $N = 104$: $r = .20$, $p = .05$ and $r = .25$, $p = .01$.
[b] For r_{pb} values t significant at .05 and .01 when $r = .18$ and .23, respectively.

administered somewhat differently. At the beginning of the semester, subjects were told that the Psychology Department was making a survey of the attitudes of some of the students toward some of the term projects, and that they might be asked to fill out a questionnaire after some of their assignments. Following participation in the Spool Packing Task the subject was thanked for his help and dismissed. Then, as if by error, the experimenter called the subject back in, reminded him of the department survey, and asked if he would not mind filling out the questionnaire privately and leaving it in a box outside the office door. The standard four questions were asked (Festinger & Carlsmith, 1959; Marlowe & Crowne, 1961); however, the subject checked an 11-point rating scale for all items. His reaction score was the sum of the four ratings, with the larger scores indicating more favorable attitudes.

Auxiliary measures. One of the behavioral measures (Spool Packing Task) has been used in previous research, so therefore it was decided to take advantage of the opportunity to determine whether similar findings would be obtained in this slightly modified situation. Hence, 83 of the initial group of subjects (43 males and 40 females) volunteered to take the Marlowe-Crowne Social Desirability (M-C *SD*) scale (Crowne & Marlowe, 1960) and the Barron (1953) Independence of Judgment Scale. This step was added after the project was initially designed and is therefore not discussed fully in this paper. The results are meaningful for this study however, and are included in Table 1 and the discussion.

RESULTS

The major analysis of the data was the intercorrelation of all the test scores and behavioral task performance scores. These intercorrelations are Pearson product-moment correlations. In addition, each of the variables was correlated with sex using the point biserial correlation. The reliability of each of the acquiescence and behavioral measures was determined. The split-half (odd-even) technique with the Spearman-Brown correction was applied to the acquiescence scales. Coefficient alpha (Cronbach, 1951) was used to measure the reliability of the Spool Packing questionnaire, and the split-half (odd-even) technique (without the Spearman-Brown correction) was applied to the critical Session II trials of the autokinetic task. Table 1 shows the reliabilities and interrelationships which resulted.

The most surprising finding in Table 1 is that none of the measures of acquiescence is related to either of the behavioral tasks. Only one of the correlations even approaches statistical significance at the .05 level. Furthermore, the two behavioral tasks are not correlated with one another. These low correlations suggest rather strongly that agreement to test items, attitude change under social influence, and expression of positive attitudes after a boring task, as they were measured here, are behaviors which have very little in common.

It is an important finding that all but one (ESP) of the measures proved to be measuring some trait reliably; apparently they are not measuring a common trait. The data in Table 1 replicate those reported earlier with several of the same measures and a different sample of subjects (McGee, 1962a). These data may thus be viewed as a cross-validation of earlier conclusions that several operational measures of response acquiescence do not necessarily measure a single "pure" response tendency. Among the six possible combinations of the SAS, ARS, F Scale, and Reversed F Scale, the amount of common variance ranges from 1% to 18%. Only three of these correlations are significant at the .05 level, and these are of little practical sig-

TABLE 2

MEANS AND STANDARD DEVIATIONS OF SCORES ON EACH VARIABLE

Variable	*N*	*M*	*SD*	Number of items
Acquiescence measures				
Social Acquiescence Scale	104	243.65	30.61	56
Agreement Response Scale	104	58.95	9.89	15
California F Scale	104	111.73	17.76	30
Reversed F Scale	104	135.22	13.46	30
Original-Reversed F	104	109.08	17.56	27
German Words	104	107.59	15.41	26
ESP Technique	104	116.21	11.07	25
Behavioral tasks				
Spool packing	104	25.86	5.76	4
Autokinetic Conformity[a]	104	162.06	44.05	—
Auxiliary measures[b]				
Marlowe-Crowne *SD* scale	83	14.16	5.49	33
Independence of Judgment Scale	83	10.28	3.13	22

[a] Raw shift ratios were transformed linearly to facilitate computation of *r* values.

[b] Responses to these measures were scored dichotomously to facilitate comparisons with previous findings.

nificance (Edwards, 1950) in the search for sources of common variance.

It is interesting to note that the Barron scale shows a significant correlation with both the original F Scale and the SAS, yet it does not correlate with the autokinetic conformity task. It does, however, correlate −.37 with the M-C *SD* scale, which is a low but significant relationship.

DISCUSSION

The data presented in Table 1 offer no support whatever for either of the two hypotheses under investigation here. One is forced to conclude that the agreeing response tendency has very little relationship, if any, to social behavior as it was measured in this study by methodologically independent, nonpsychometric tasks. From observing statistically significant correlation coefficients between biased measures one may well conclude that response acquiescence is related to conformity. The correlation between the SAS and the Barron scale would appear on the surface to offer support for the conclusion reached by Bass (1956) that a person with a high social acquiescence score is ". . . in short, a Babbitt—an unquestioning conformer to social demands placed on him" (p. 297). Similarly, the correlation between the Barron scale and the original F Scale would tend to confirm Jackson's (1958, 1959) findings of an indirect relationship between acquiescence and social independence. However, when conformity is measured behaviorally within the context of an actual social pressure situation, there is no relationship between the two variables.

It is noteworthy that the Barron (1953) scale did not predict conformity in the autokinetic task. However, it should be remembered that the situation developed by Asch (1956) which was used to construct the Barron scale differs from the autokinetic experiment in two major respects. In the first place there is reported to be a fairly high degree of veridicality to the length-of-lines judgment task; subjects have reason to believe they are actually perceiving correctly. Yielding under these conditions may be considered a markedly different type of behavior than shifting the distance-of-movement judgment in the ambiguous autokinetic experiment. Second, in the former studies the subject was subjected to a "unanimous majority" opinion when he was a "minority of one." In the present study he was subjected to the influence of only one other person in making his verbal report.

The attempt to relate test taking response styles to underlying personality syndromes has recently received enthusiastic attention. The primary purpose of the present investigation was to demonstrate a major pitfall to be avoided in this effort. It has been shown how significant correlations between two psychometric instruments may suggest a relationship between acquiescence and conformity, or anal resentment, or extroversion, etc. These between-test correlations may be replicated from sample to sample. However, in the case of conformity, when the dependent variable is measured behaviorally, the relationship disappears.

The present study offers empirical evidence for some speculative arguments elaborated previously (McGee, 1962b). Response styles are apparently very elusive traits, if in fact they even exist in a form deserving of the designation "trait" (McGee, 1962a). Hence, they are quite difficult to measure reliably, and may have quite questionable validity for predicting personality traits upon which they are assumed to be based. *There is still a need for evidence to support this popular assumption.*

The very term "acquiescence" has many and varied meanings. For some investigators it refers to a specific test taking behavior, i.e., *the number of "agree" responses, irrespective of content.* For others the term denotes a nebulous construct—a social behavior within a family of traits including conformity, persuasability, suggestibility, and submission to authority. It appears that recent researchers have failed to make any clear operational distinction between what might be called "social acquiescence" and "response acquiescence."

SUMMARY

The present study was designed to assess the currently popular hypothesis that measures of

response sets, such as acquiescence, may be used to predict certain personality traits with which they are assumed to be dynamically related. To avoid the possibility of biasing the data with methodological artifacts, nonpsychometric behavioral tasks were employed to measure the dependent variables. In this study, conformity and tendency to seek social acceptance were used as the behavioral measures. Seven different scales were used to measure the agreeing response tendency. The results showed clearly that while acquiescence scores tend to correlate with other psychometric instruments, they do not correlate with the behavioral measures of social responses used here. The present evidence continues to indicate that this popular hypothesis is still in need of empirical support.

REFERENCES

ADORNO, T. W., FRENKEL-BRUNSWIK, ELSE, LEVINSON, D. J., & SANFORD, R. N. *The authoritarian personality*. New York: Harper, 1950.

ASCH, S. E. Studies of independence and submission to group pressure: I. A minority of one against a unanimous majority. *Psychol. Monogr.*, 1956, **70**(9, Whole No. 416).

BARRON, F. Some personality correlates of independence of judgement. *J. Pers.*, 1953, **21**, 287–297.

BASS, B. M. Development and evaluation of a scale for measuring social acquiescence. *J. abnorm. soc. Psychol.*, 1956, **53**, 296–299.

CHRISTIE, R., HAVEL, JOAN, & SEIDENBERG, B. Is the F Scale irreversible? *J. abnorm. soc. Psychol.*, 1958, **56**, 143–159.

COUCH, A., & KENISTON, K. Yeasayers and naysayers: Agreeing response set as a personality variable. *J. abnorm. soc. Psychol.*, 1960, **60**, 151–174.

CRONBACH, L. J. Coefficient alpha and the internal structure of tests. *Psychometrika*, 1951, **16**, 297–334.

CROWNE, D. P., & MARLOWE, D. A new scale of social desirability independent of psychopathology. *J. consult. Psychol.*, 1960, **24**, 349–354.

CROWNE, D. P., & STRICKLAND, BONNIE R. The conditioning of verbal behavior as a function of the need for social approval. *J abnorm. soc. Psychol.*, 1961, **63**, 395–401.

EDWARDS, A. L. *Experimental design in psychological research*. New York: Rinehart, 1950.

FESTINGER, L., & CARLSMITH, J. M. Cognitive consequences of forced complicance. *J. abnorm. soc. Psychol.*, 1959, **58**, 203–210.

GAGE, N. L., LEAVITT, G. S., & STONE, G. C. The psychological meaning of acquiescence set for authoritarianism. *J. abnorm. soc. Psychol.*, 1957, **55**, 98–103.

GRAY, C. W., & CRISP, H. E. The credibility of pure response set. Paper read at Southeastern Psychological Association, Gatlinburg, April 1961.

JACKSON, D. N. Independence and resistance to perceptual field forces. *J. abnorm. soc. Psychol.*, 1958, **56**, 279–281.

JACKSON, D. N. Cognitive energy level, response acquiescence, and authoritarianism. *J. soc. Psychol.*, 1959, **49**, 65–69.

MCGEE, R. K. The relationship between response style and personality variables: I. The measurement of response acquiescence. *J. abnorm. soc. Psychol.*, 1962, **64**, 229–233. (a)

MCGEE, R. K. Response style as a personality variable: By what criterion? *Psychol. Bull.*, 1962, **59**, 284–295. (b)

MARLOWE, D. Need for social approval and the operant conditioning of meaningful verbal behavior. *J. consult. Psychol.*, 1962, **26**, 79–83.

MARLOWE, D., & CROWNE, D. P. Social desirability and response to perceived situational demands. *J. consult. Psychol.*, 1961, **25**, 109–115.

SHERIF, M. A study of some social factors in perception. *Arch. Psychol., N. Y.*, 1935, No. 187.

STRICKLAND, BONNIE R., & CROWNE, D. P. Conformity under conditions of simulated group pressure as a function of the need for social approval. *J. soc. Psychol.*, in press.

(Received July 17, 1961)

Journal of Abnormal and Social Psychology
1962, Vol. 65, No. 5, 351–354

THE INSTRUMENTAL ESCAPE CONDITIONING OF ANXIOUS AND NONANXIOUS CHILDREN [1]

R. K. PENNEY AND B. MCCANN

University of Texas

There have been several investigations concerned with the relationship between stimulus intensity and anxiety. Wenar (1954), using college subjects, determined the effect of level of anxiety and level of stimulus intensity upon reaction time. The Taylor Manifest Anxiety (*MA*) scale was used and subjects whose scores were in the upper 20% of the scale were defined as anxious and those in the lower 20% as nonanxious. Two in-

[1] This research was performed while both authors were at McMaster University, Hamilton, Ontario. The authors wish to express their sincere thanks to George Walker and his staff at Strathcona School for their cooperation.

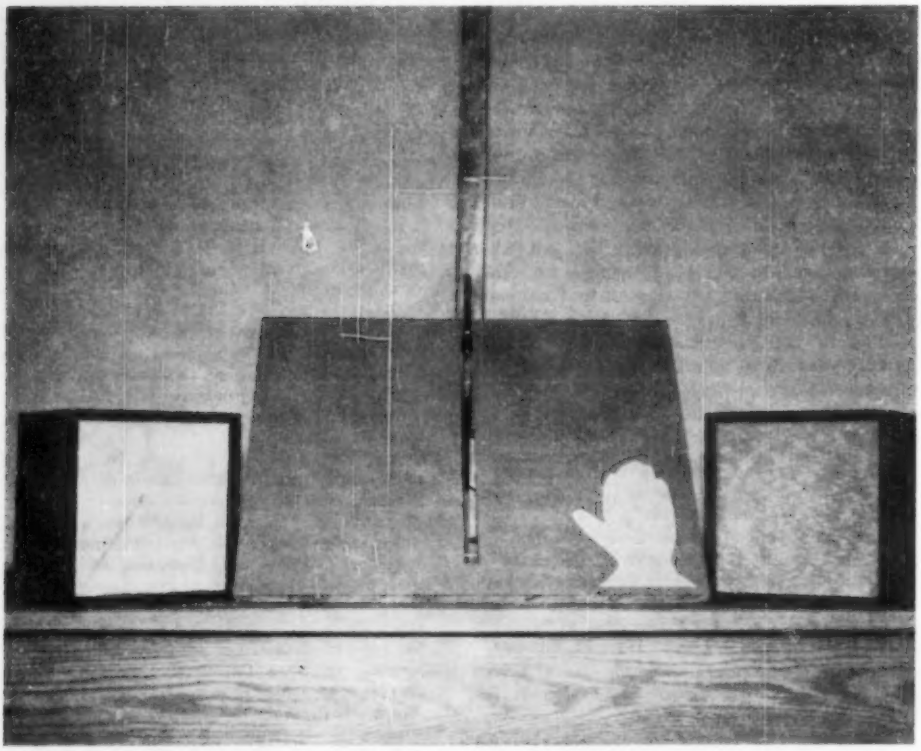

FIG. 1. Experimental apparatus.

tensities of electric shock were employed as stimuli. The results of Wenar's study indicated that an increase in stimulus intensity, i.e., a stronger electric shock, significantly increased the speed of reaction for both anxious and nonanxious subjects. Furthermore, the anxious group responded faster than the nonanxious group at both levels of shock intensity with the difference between the two groups remaining relatively constant.

Castaneda (1956) investigated the effects of two levels of auditory stimulus upon the reaction time and response amplitude of anxious and nonanxious high school subjects. The *MA* was employed to define anxious and nonanxious subjects, and a mild and a raucous buzzer constituted the stimuli. The response element was a hand dynamometer. Castaneda found a significant interaction between anxiety and stimulus intensity with the anxious group responding slower than the nonanxious group at the weak intensity (mild buzzer) and faster at the high intensity (raucous buzzer). Mean amplitude scores for the anxious group were significantly greater than those of the nonanxious group at both levels of stimulus intensity.

The present study used elementary school children as subjects and was designed to determine the relationship between stimulus intensity and level of anxiety. It differs from those studies cited above, however, in that subjects were placed in an instrumental escape situation and were not told that the speed of their response was being measured.

METHOD

Subjects

Sixty-nine children in Grade 3 and 32 in Grade 4 at Strathcona Public School, Hamilton, Ontario, were administered Sarason's (Sarason, Davidson, Lighthall, Waite, & Ruebush, 1960) Test Anxiety Scale for Children (TASC) and the General Anxiety Scale for Children (GASC). Children with lie scores greater than 6 were discarded. The TASC and GASC scores were combined and an anxious and a nonanxious group

consisting of 22 children each were formed. The mean total anxiety score for the anxious group was 46.91 with a range of 37–55. The minimum TASC score for the anxious group was 11 while the minimum GASC score was 22. The mean total anxiety score for the nonanxious group was 18.77 with a range of 3–27. The maximum scores for the nonanxious group were 10 on the TASC, and 18 on the GASC.

Apparatus

The apparatus consisted of a wooden box (Figure 1) painted flat gray. The response element, a lever, projected from a sloping panel which faced the subject. This front panel was 18 inches × 15 inches and set at an angle of 45° with the table upon which the apparatus was placed. The lever was in the middle of the sloping panel, and rested at the top of the panel in the normal position. The lever moved toward the subject through an excursion of 12 inches. On the face of the panel to the lower right of the lever was a white hand pattern upon which the subject rested his hand between trials.

An Ashman tone generator emitted tones of 98 and 65 db. at 2000 cps as measured by a General Sound Survey meter. The tones acted as signals for the subject to pull the lever and were terminated when the lever reached the end of its excursion.

When the subject pulled the lever away from the normal resting position, a Standard Electric timer was activated. The timer was deactivated at the completion of the excursion. When the subject released the lever, it was returned to its normal position by means of a spiral spring.

Procedure

The subjects were sent individually from the classroom to a trailer parked on the school grounds. The experimenter introduced himself to the child and read the following instructions:

> Place your hand here [hand pattern was indicated] and when you hear a tone pull the lever all the way to the bottom, and then put your hand back on the pattern. Okay? I will be in the next room. Remember, wait until you hear the tone before you pull the lever.

After reading the instructions, the experimenter retired to an observation booth at the back of the trailer from which he could view the subject through a one-way mirror.

Eleven anxious and 11 nonanxious subjects were given 20 presentations of the 98-db. tone. Similarly, subjects were given 20 presentations of the 65-db. tone. The time interval between trials was not constant but varied with a maximum intertrial interval of 10 seconds.

RESULTS

Figure 2 presents the mean response time for the four groups over the five blocks of four trials each. An analysis of variance was performed over these data (Lindquist, 1953, Type III). This

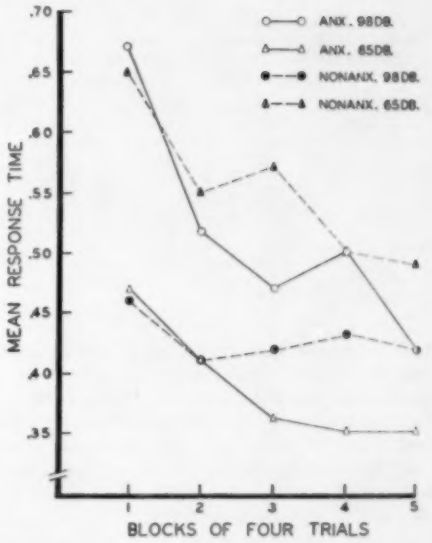

FIG. 2. Mean response time for all four groups over the five blocks of four trials.

analysis revealed a significant ($p < .001$) triple interaction (Trials × Anxiety × Stimulus Intensity) indicating that the difference between the anxious and nonanxious groups depended on the level of stimulus intensity and the block of trials. Accordingly, two additional analyses were conducted. The first analysis (Table 1) compared the anxious and nonanxious groups at the 65-db. level of intensity. A significant main effect for anxiety was found, i.e., the anxious group responded significantly faster than the nonanxious group over all blocks of trials. The second analysis (Table 2) compared the anxious and nonanxious groups at the 98-db. level of stimulus intensity. This analysis yielded a significant inter-

TABLE 1

SUMMARY OF ANALYSIS OF VARIANCE FOR ANXIOUS AND NONANXIOUS CHILDREN RESPONDING UNDER THE 65-DB. LEVEL OF INTENSITY

Source	*df*	*MS*	*F*
Anxiety (A)	1	.71	7.10*
Error (b)	20	.10	
Trials (T)	4	.07	11.67***
T × A	4	.007	1.17
Error (w)	80	.006	

* $p < .025$.
*** $p < .091$.

TABLE 2

SUMMARY OF ANALYSIS OF VARIANCE FOR ANXIOUS AND NONANXIOUS CHILDREN RESPONDING UNDER THE 98-DB. LEVEL OF INTENSITY

Source	*df*	*MS*	*F*
Anxiety (A)	1	.22	1.79
Error (b)	20	.123	
Trials (T)	4	.07	—
T × A	4	.035	3.89**
Error (w)	80	.009	

** $p < .01$.

action between trials and anxiety and indicated that the difference between the two groups was a function of the block of trials. Using Cochran's approximate t test, the differences were tested at each block of trials. The only significant difference occurred at the first block of trials where the anxious group was significantly ($p < .05$) slower than the nonanxious group.

DISCUSSION

According to Spence (1956) anxiety acts as an irrelevant drive and serves to energize all existing habits. Thus, an anxious group would have greater drive strength than a nonanxious group. As the theory would predict, the anxious group in the present study, responded faster than the nonanxious group under the 65-db. level of intensity. However, a reversal of these results was found for subjects responding under the 98-db. tone. The theory would predict this result if the 98-db. tone introduced suprathreshold responses which under the high drive condition successfully competed with the lever pulling response. These same competing responses would not be suprathreshold (or to a lesser degree above threshold) under the nonanxious condition due to the lower drive strength. Furthermore, according to this hypothesis, as the lever pulling response gains in strength over trials, and/or as the competing responses adapt out, the differences between the anxious and nonanxious groups should disappear. The data supports this interpretation although there is no independent evidence for the presence of competing responses in either group.

Assuming that the competing responses adapt out, one inference is that if training were extended under the 98-db. condition, the anxious group would at some point respond faster than the nonanxious group. Unfortunately, only 20 trials were explored.

SUMMARY

Forty-four elementary school children were administered Sarason's anxiety scales and a combined score yielded anxious and nonanxious groups. Eleven anxious and 11 nonanxious subjects were given 20 presentations of a 65-db. tone. Similarly, 11 anxious and 11 nonanxious subjects were given 20 presentations of a 98-db. tone. The subjects could escape the tones if they pulled a lever through its excursion.

The anxious subjects responding to the 65-db. tone lever pulled significantly faster than the nonanxious subjects. However, under the 98-db. condition, the nonanxious subjects responded faster than the anxious subjects although this difference was a function of trials. These results were interpreted in terms of the interaction theory of Hull-Spence.

REFERENCES

CASTANEDA, A. Reaction time and response amplitude as a function of anxiety and stimulus intensity. *J. abnorm. soc. Psychol.*, 1956, **53**, 225–228.

LINDQUIST, E. F. *Design and analysis of experiments in psychology and education*. Cambridge: Houghton Mifflin, 1953.

SARASON, S., DAVIDSON, K., LIGHTHALL, F., WAITE, R., & RUEBUSH, B. *Anxiety in elementary school children*. New York: Wiley, 1960.

SPENCE, K. W. *Behavior theory and conditioning*. New Haven: Yale Univer. Press, 1956.

WENAR, C. Reaction time as a function of manifest anxiety and stimulus intensity. *J. abnorm. soc. Psychol.*, 1954, **49**, 335–340.

(Received July 17, 1961)

Volume II

Hans H. Strupp
Lester Luborsky
Editors

of

RESEARCH in PSYCHOTHERAPY

Proceedings of the Second Conference on Research in Psychotherapy held at Chapel Hill, N.C., in May 1961. Leading research workers in psychology and psychiatry deal with selected research issues and assess new research developments. Their efforts place research in psychotherapy more squarely within psychological research concerned with understanding interpersonal processes.

The psychotherapist's contribution to the treatment process

Measuring personality change in psychotherapy

Definition, measurement, and analysis of significant variables

The book is a follow up of *Research in Psychotherapy*, published by the American Psychological Association in 1959, which reported on the First Conference on Research in Psychotherapy held in Washington, D.C., in 1958.

Contributors

Barbara J. Betz
Jack Block
Edward S. Bordin
John M. Butler
Kenneth Mark Colby
Jerome D. Frank
Ernest A. Haggard
Leonard Krasner
Henry L. Lennard
Daniel J. Levinson
Maurice Lorr
Lester Luborsky
Morris B. Parloff
Laura N. Rice
Irwin C. Rosen
Nevitt Sanford
George Saslow
David Shakow
John M. Shlien
Richard S. Siegal
William U. Snyder
Ian Stevenson
Hans H. Strupp
Alice K. Wagstaff

329 pp.
plus Indexes
$3.50

Order from:
AMERICAN PSYCHOLOGICAL ASSOCIATION
1333-16th St., N.W., Washington 6, D.C.

"The new in tests. . . . . ."

Need an individual test which *quickly* provides a stable and reliable estimate of intelligence in 2 or 3 minutes per form? Has three forms? Does not depend on verbal production or subjective scoring? Can be used with the severely physically handicapped (even paralyzed), if they can signal yes-no? Handles two-year-olds and superior adults within the same short series of items and the same format? Has been painstakingly standardized? Builds rapport with intrinsically interesting and simple procedures based on multiple drawings?

Ammons and Ammons

QUICK TEST (QT)

Order Tester's Set including manual, card materials, keys, and 100 answer sheets ($8.00 complete).

For further information, write to

PSYCHOLOGICAL TEST SPECIALISTS
Box 1441
Missoula, Montana

"The new in tests. . . . . ."